AD-A206 947

DTIC ELECTE APR 13 1989 S H D

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

UNCLASSIFIED

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release, distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 89-0402
6a. NAME OF PERFORMING ORGANIZATION: Oregon State University Electrical & Computer Engr.	6b. OFFICE SYMBOL (If applicable)
6c. ADDRESS (City, State and ZIP Code): Corvallis, OR 97331	7a. NAME OF MONITORING ORGANIZATION: AFOSR/NE
	7b. ADDRESS (City, State and ZIP Code): Bldg. 410, Bolling AFB, DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR	8b. OFFICE SYMBOL (If applicable): NE
8c. ADDRESS (City, State and ZIP Code): Bldg. 410, Bolling Air Force Base, DC 20332	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-86-0309
	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2306	B1	

11. TITLE (Include Security Classification): Point Defects in Semiconductors: Microscopic Identification.

12. PERSONAL AUTHOR(S): Professor James A. Van Vechten and Dr. John F. Wager

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final	FROM 31/8/86 TO 23/2/89	1989 March 31	

16. SUPPLEMENTARY NOTATION

17 COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The goal of the research program described herein was to provide insight into the identity and properties of point defects in semiconductors. Particular emphasis was devoted to problems involving microscopic identification, metastable properties, defect migration, and diffusion of point defects in semiconductors. Our approach was to apply atomistic thermodynamic theory, Monte Carlo simulation, and experimental analysis to elucidate the nature and properties of semiconductor defects. Significant progress has been made in the following seven areas: 1) recombination enhanced vacancy migration in silicon, 2) Monte Carlo simulation of diffusion in semiconductors, 3) phosphorous vacancy nearest-neighbor hopping in InP, 4) entropy of migration for atomic hopping, 5) EL2/ELO identification in GaAs, 6) characterization and identification of DX in AlGaAs, and 7) temperature dependence of band offsets.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☒ DTIC USERS ☐	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
LT. COL. KELTON	(202) 767-4931	NE

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE

CONT'D FROM #11.

METASTABLE PROPERTIES, DEFECT MIGRATION, AND DIFFUSION

FINAL TECHNICAL REPORT

to

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

BOLLING AIR FORCE BASE
WASHINGTON, D.C. 20332

on

POINT DEFECTS IN SEMICONDUCTORS: MICROSCOPIC IDENTIFICATION, METASTABLE PROPERTIES, DEFECT MIGRATION, AND DIFFUSION

(AFOSR-86-0309)

For Period

31 August 1986 to 31 March 1989

Submitted by

Professor James A. Van Vechten

and

Assistant Professor John F. Wager

Department of Electrical & Computer Engineering
Oregon State University
Corvallis, OR 97331

ABSTRACT

The goal of the research program described herein was to provide insight into the identity and properties of point defects in semiconductors. Particular emphasis was devoted to problems involving microscopic identification, metastable properties, defect migration, and diffusion of point defects in semiconductors. Our approach was to apply atomistic thermodynamic theory, Monte Carlo simulation, and experimental analysis to elucidate the nature and properties of semiconductor defects. Significant progress has been made in the following seven areas:

1. Recombination enhanced vacancy migration in silicon,
2. Monte Carlo simulation of diffusion in semiconductors,
3. Phosphorous vacancy nearest-neighbor hopping in InP,
4. Entropy of migration for atomic hopping,
5. EL2/ELO identification in GaAs,
6. Characterization and identification of DX in AlGaAs, and
7. Temperature dependence of band offsets.

A summary of this work is as follows, while a detailed discussion is available in the publications enclosed with this report.

1. Recombination Enhanced Vacancy Migration: "Activation of Enthalpy of Recombination-Enhanced Vacancy Migration in Si," J.A. Van Vechten, *Phys. Rev. B. 38*, 9913 (1988).

Much of our work on defect dynamics is based on the validity of the Ballistic Model (BM) for vacancy migration which asserts (essentially) that the activation barrier for vacancy migration ΔH_m, is determined by the kinetic energy of the atom which hops into the vacancy. One assumption implicit in the BM is that the temperature exceeds the Debye temperature of the material

in question. In the above paper, the BM is extended to temperatures below the Debye temperature. The intent of this work is to explain for the first time experimentally determined ΔH_m's for various charge states in terms of ΔH_m inferred from the BM reduced by the appropriate enthalpy of recombination, ΔH_R, to give the experimentally observed enthalpy of recombination enhanced vacancy migration, ΔH_{REVM}. The significant results are: (a) a thermal REVM is explained as a natural consequence of $\Delta H_R > \Delta H_m$ for all charge states which have been determined experimentally, and (b) at high temperatures thermal REVM, TREVM, makes a significant contribution to the total vacancy migration since thermally generated free carriers are present in high concentrations and recombine at vacancies; TREVM is proposed as the reason for the curvature of experimental Arrhenius plots for self-diffusion.

2. **Monte Carlo Simulation of Diffusion in Semiconductors: "Vacancy First and Second Neighbor Hopping at a Compound Semiconductor Interfaces – Insights from Computer Simulation," J.A. Van Vechten and U. Schmid, *J. Vac. Sci. Technol.* (submitted); "Diffusion of Au into Si Via Simple Frank-Turnbull Mechanism and Kick-Out Mechanism Examined by VIDSIM Computer Simulation," U. Schmid and J.A. Van Vechten (in preparation).**

VIDSIM consists of 7000 lines of 'C' language code and the user inputs whatever assumptions he chooses for the interaction parameters among host and impurity atoms at lattice sites, antisites, bond-centered interstitial sites, hexagonal interstitial sites or tetrahedral interstitial sites and vacant sites. The user also inputs an initial configuration of the point defects in the crystal, the temperature and the "seed" for a random number generator. The program then uses a Monte Carlo method to work out the consequent migration and reaction of the point defects through many discrete events. At present, the program is limited to diamond or zinc-blende structure lattices of no more than 2^{42} (4.4×10^{12}) sites with no more than 200 mobile defects

(vacancies and interstitials) and about 5000 other defects (impurities and antisites) that move only by interaction with the mobile defects. When run on an IBM PC/AT with a 6 Mhz math coprocessor and 2 nn hopping switched off, it typically simulates 50 discrete events (e.g., vacancy hop to first neighbor site, interstitial hop, vacancy-interstitial creation or annihilation) per second. When we allow 2 nn hopping, the rate slows to about 14 events per second. The rate on other computers scales very closely with the frequency of the math coprocessor.

It should cost the user almost nothing to use VIDSIM. The program is designed to run in "background" with commonly available PC multitasking utilities such as SoftLogic's DoubleDos or Quaterdeck's DesqView while the computer is used for other purposes or while it would otherwise stand idle over night, etc. It requires only 225 Kbytes of the computer's RAM memory for the simulation described here. A small simulation requires as little as 120 Kbytes.

We have studied the Al diffusion in AlGaAs with the aid of simulation, VIDSIM, that works out the complex consequences of ones assumptions re thermochemical parameters at the atomic level. Our previously published BM assumptions lead to the conclusion that Si promotes the conversion of Ga vacancies to trivacancies while Te does not; they also identify 2.9 eV as the sum of the Ga second neighbor hopping enthalpy and the enthalpy of formation of single negative Ga vacancy in n-type GaAs and identify 4 eV as the sum of the enthalpy of formation of the trivacancy plus its nearest neighbor hopping energy in GaAs. The simulation also finds that Al by itself converts As vacancies to trivacancies plus metal antisite. It shows the trivacancies to diffuse very rapidly and to entrain Al. We propose that the cation sublattice dominance of Si induced host interdiffusion follows from this entrainment. We

also suggest these complexes are related to DX and account for the anomalous variation of the parameters of that defect with Al concentration.

One of the first problems VIDSIM was used to simulate was the diffusion of Au into Si according both to the assumptions of the Frank-Turnbull (vacancy + Au interstitial) mechanism and the "kick-out" (host interstitial + Au interstitial) mechanism. We have shown that the advocates of the kick-out mechanism erred in their efforts to deduce the macroscopic consequences of their assumptions regarding the atom level processes. In particular, we have identified a central error in their attempt to model the diffusion processes by analytic methods (approximate solutions of differential equations with assumptions of boundary conditions); they assume that the Au interstitial concentration becomes uniform throughout the sample early in the diffusion process while the substitutional Au is confined near its source. We have shown that the assumption that the Au interstitials can displace Si atoms from the lattice, the central tenant of the kick-out model, leads to an essentially exponential decay of Au interstitial concentration into regions free of Au substitutionals. We have further shown that no assumption about the rate of diffusion or interaction of the Si self-interstitials, which are hypothesized in the kick-out model, can significantly alter this feature of the Au interstitial profile. We have shown that, while the advocates of the kick-out mechanism claim it leads to a concentration of Au substitutionals that varies as the square root of time, the model in fact leads to a strictly linear variation of the Au substitutional concentration at all depths and for all times. The advocates of the kick-out mechanism also claimed that, contrary to the claims of the advocates of the Frank-Turnbull mechanism, the Frank-Turnbull mechanism must lead to an Au substitutional profile which is a complementary error function of depth and linear in time. Our simulation found that,

for parameters previously estimated for the Frank-Turnbull mechanism, the simulated profile did have these features for quite a long time of simulation (about 4×10^8 events). This resulted from the uninhibited diffusion of the Au interstitials throughout the sample early in the simulation when there were appreciable vacancy concentrations only near the surfaces. (In the Frank-Turnbull model Au cannot displace Si from the lattice so the Au diffuses rapidly and without forming substitutionals where there are no vacancies.) However, at later times in the simulation, a sufficient concentration of vacancies diffused into a sufficient depth that the new Au interstitials diffusing in from their source at the surface had a high probability to react with one of these vacancies and become a substitutional. This throttled down the supply of Au to the interior of the sample and caused a distinct change in both the profile with depth and the rate of increase with time. As the simulation developed further, the profile became much more steeply peaked at the surface than an erfc(z) and the rate of increase turned to something well approximated by a square root of time variation.

3. Phosphorous Vacancy Nearest-Neighbor Hopping: "Phosphorous Vacancy Nearest-Neighbor Hopping Involved Instabilities in InP Capacitors," M.T. Juang, J.F. Wager, and J.A. Van Vechten, *J. Electrochem. Soc. 135*, 2019, 2023 (1988).

We regard this work as strong experimental support for the validity of the BM. In 1985 we (J.A. Van Vechten and J.F. Wager, *J. Appl. Phys. 57*, 1956 (1985)) offered an explanation for drain current drift in InP MISFETs. This explanation involved a Fermi-level-induced migration of an In atom onto a nearest-neighbor phosphorous vacancy (thus, phosphorous vacancy nearest-neighbor hopping (PVNNH)) and a concomitant capture of channel electrons resulting in a decrease in drain current. We predicted an activation

enthalpy, ΔH_m = 1.15 eV, which is in excellent agreement with the experimentally deduced values of 1.1 to 1.2 eV reported in the above papers from variable-temperature bias-stress measurements. The significant results are: (a) ΔH_m found experimentally is in excellent agreement with that predicted theoretically, (b) the agreement between experiment and theory we regard as strong support for the importance of nearest-neighbor hopping as a diffusion mechanism in semiconductors, (c) a computer simulation of the experimental data enabled us to determine the rate-limiting step of the PVNNH defect reaction, and (3) the computer simulation required that the activation entropy of the rate-limiting step be chosen as ΔS_m = 15.3 k; ΔS_m was the only adjustable parameter in the simulation. (Points (c) and (d) are important with respect to the following issue regarding the entropy of migration.)

4. Entropy of Migration: "Entropy of Migration for Atomic Hopping," T.W. Dobson, J.F. Wager, and J.A. Van Vechten (submitted to *Phys. Rev. B*).

In our PVNNH work we deduced an abnormally large entropy of migration, ΔS_m = 15.3 k. Although this value of ΔS_m is "abnormal" in the sense that there is currently no explanation available in the literature for ΔS_m's of this magnitude, such values have frequently been reported experimentally in a variety of solid-state processes such as self-diffusion and defect transformations. In the paper referenced above, we have employed the BM hypothesis that atomic migration is kinetic in origin. In the spirit of the BM, we propose that ΔS_m is the translational entropy of the hopping atom in its addle point configuration minus the vibrational entropy of the hopping atom in its equilibrium position prior to the hop. These translational and vibrational entropies are calculated using standard statistical thermodynamic expressions, assuming that the translational entropy is given by the particle in a box

approximation and the vibrational entropy is that of a simple harmonic oscillator vibrating at the Debye frequency. This formulation is then quantitatively applied to explain the abnormal prefactors experimentally deduced in PVNNH measurements, DLTS studies of the metastable M center in InP, and Si and Ge self-diffusion experiments.

5. EL2/EL0 in GaAs: "Atomic Model for the EL0 Defect in GaAs," J.F. Wager and J.A. Van Vechten, *J. Appl. Phys. 62*, 4192 (1987); "Reply to: 'Comment on 'Atomic Model for the EL2 Defect in GaAs'' by H.J. Von Bardeleben, J.C. Bourgoin, and D. Stievenard," J.F. Wager and J.A. Van Vechten, *Phys. Rev. B 39*, 1967 (1989); "Reply to: 'Comment on 'Atomic Model for the EL2 Defect in GaAs'' by G. Wang, Y. Zou, S. Benakki, A. Goltzene, and C. Schwab," *Phys. Rev. B 38*, 10956 (1988).

In 1987 we proposed (J.F. Wager and J.A. Van Vechten, *Phys. Rev. B 35*, 2330 (1987)) atomic models for the stable and metastable configurations of EL2 in GaAs and identified the transformation mechanism between configurations as nearest-neighbor hopping. In the first paper referenced above, we propose similar atomic models for the bi-stable, oxygen-related defect in GaAs known as EL0. In the stable configuration, we propose that EL2 and EL0 consist of divacancies with a neighboring As_{Ga} or O_{As}, respectively. In the metastable state, the vacancies comprising the defect complex have been physically separated as a result of atomic rearrangement by nearest-neighbor hopping. We remain convinced of the essential validity of our atomic identifications of EL2 and EL0 and have defended our model from an attack by the authors listed in the second reference above who have proposed an alternative model for EL2 consisting of an As_{Ga} antisite and an As interstitial. Additionally, we have presented a thermodynamic argument, using these authors' own data, which reveals one of the fatal flaws in their atomic model.

The primary issue raised by Wang et al. in their comment reference above regards the geometry of our proposed atomic model, not its components. They

favor a geometry for the stable configuration in which As_{Ga} occupies a nearest-neighbor site to the V_{As} portion of the divacancy while we originally proposed As_{Ga} to be a next-nearest-neighbor to the V_{Ga} component of the divacancy. Our original identification was predicated by EPR and ENDOR measurements which has always required As_{Ga} to have four As nearest-neighbors. Wang et al. are now questioning this interpretation of the EPR and ENDOR experiments. Without becoming immersed in a discussion of how to best interpret the EPR and ENDOR experiments, if we tentatively accept the revised geometry suggested by Wang et al., we can then argue (as we did in the third paper referenced above), that the most sensible identification of the metastable configuration of EL2 is that of two V_{Ga}'s which arise from a single nearest-neighbor hop of As_{Ga} in the stable state. This interpretation is appealing in light of recent work by Mitchel, Rea, and Yu (submitted to *J. Appl. Phys.*) in which they found type conversion from n-type to p-type in EL2-dominated GaAs samples upon transformation to the metastable state. We have proposed to Phil Yu, WPAFB, that photocapacitance quenching (PCQ) experiments of EL2- and ELO-dominated samples should provide an important test for identification of the geometry of EL2, the metastable state of EL2, and the transformation mechanism from the stable to metastable configuration.

In summary: (a) we stand by our original identification of EL2 and ELO as divacancies with neighboring As_{Ga} and O_{As}, respectively; (b) we are leaning towards the revised geometry of the stable configuration of EL2 as proposed by Wang et al.; (c) we propose a revised model for the metastable configuration of EL2 consisting of two V_{Ga}'s; (d) we propose that the PCQ experiments of P. Yu should establish the validity of the revised geometry and metastable configuration; and (e) we point out that much of the mystery regarding EL2 is that GaAs samples contain varying amounts of EL2 and ELO and that although

these defects have very similar physical properties, there are subtle differences which have led to much confusion in the microscopic identification of EL2.

6. DX in AlGaAs: (a) "Mössbauer Spectra and the DX Center Complex in AlGaAs," J.A. Van Vechten (accepted by *J. Phys. Condensed Matter Physics*).

Mössbauer spectra for the "DX" deep level defect in AlGaAs alloys are just now being made available, e.g., by Gibart et al. (1988). The interpretation of these spectra has been somewhat clouded by the fact that Mössbauer specialists have generally not considered the antisite defect among the defect complexes present in III-V crystals near the Mössbauer isotope that affect the various components of the spectra. We note that the antisite defect is well established as a component of, e.g., the EL2 deep level defect, and generally as the most prevalent native point defect in III-V's. Furthermore, the spectra of Gibart et al. are consistent with the model that the "X" of the DX defect complex is the metal on As site antisite defect plus Ga vacancy pair that results from As vacancy nearest-neighbor hopping. In particular, they show that a major fraction of donors remain in an unperturbed state when DX compensates the sample. The same argument and conclusion follow if one supposes that the "X" is a three vacancy complex rather than a single vacancy. Van Vechten and Schmid (see #2) found such three vacancy complexes formed in their Monte Carlo simulation and proposed that they would resolve mysteries re the distribution of Al around the DX center and the failure of extrapolation of DX ionization energies to pure GaAs to account for the resonant DX level found there. It would also account for the observation of neutral donor levels found by TDPPC experiments of Dobson and Wager (see #7).

(b) "Transient Decay of Persistent Photoconductance in AlGaAs," T.W. Dobson and J.F. Wager (in preparation).

We have recently completed our first phase of transient decay of persistent photoconductivity (TDPPC) measurements in which we monitor the 1 MHz small signal conductance subsequent to photon illumination. These measurements are performed over a wide range of temperature (50 to 150 K) and photon energy (0.1 to 1.8 eV) TDPPC curves are analyzed by computer simulation assuming that TDPPC is the result of Shockley-Read-Hall (SRH) recombination into photoionized deep levels which possess a capture barrier and that the carrier mobility changes concomitant with electron capture because of a change in the electron capture barrier for DX consistent with other reports in the literature, (b) we conclude that DX is donor-like which is consistent with Nelson and Lang et al. but inconsistent with Chand et al. and Saxena, and (c) we conclude that photo-Hall measurements are not as reliable as TDPPC measurements for determining the acceptor- or donor-like nature of DX because they are strongly affected by the multilayer nature of the AlGaAs experimental structures.

7. Temperature Dependence of Band Offsets: "Temperature Dependence of the Mercury Telluride–Cadmium Telluride Band Offset," K.J. Malloy and J.A. Van Vechten, *Appl. Phys. Lett. 54*, 937 (1989). "Temperature Dependence of Band Offsets for Heterojunctions in General and for the Particular Cases of AlAs–GaAs and HgTe–CdTe," J.A. Van Vechten and K.J. Malloy, *Phys. Rev. B* (accepted).

A simple thermodynamic theory is devised for the temperature dependence of the valence and conduction band offsets in semiconductor heterojunctions. It is shown how this temperature dependence is related to the bandgap temperature variations of the two substances at the junction. This theory is then employed to reconcile discrepancies among various researchers in the assessment of band offsets for the AlAs–GaAs and HgTe–CdTe materials systems.

Activation enthalpy of recombination-enhanced vacancy migration in Si

J. A. Van Vechten
Center for Advanced Materials Research, Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon 97331
(Received 19 February 1988)

As reported by Watkins and co-workers, recombination-enhanced migration is observed in Si for positively charged vacancies, but not for negatively charged vacancies. Although there is a component of the recombination-enhanced vacancy migration that is independent of temperature (athermal), we show that there is another component which obeys a simple Arrhenius behavior with an activation enthalpy of 0.07 eV. These facts are rather mysterious if one believes that the normal thermal activation enthalpies for vacancy migration measured at the temperatures used in these experiments are relevant parameters for the recombination-enhanced migration process. We show that if, instead, one accepts as relevant parameters the ballistic model values for the enthalpy of vacancy migration at high temperatures, then a simple explanation of the facts is readily apparent. At high temperatures, thermally generated carriers recombine at vacancies even in the absence of electrical or optical injection. The implications of this on the temperature dependence of the vacancy contribution to self-diffusion are discussed.

Watkins and co-workers have inferred[1] values for the activation enthalpy for normal thermal migration ΔH_m of the single vacancy V in Si in three of its five charge states as follows:

$$\Delta H_m(V^{2-})=0.18\pm0.02 \text{ eV}, \quad (1)$$

$$\Delta H_m(V^{0})=0.45\pm0.04 \text{ eV}, \quad (2)$$

$$\Delta H_m(V^{2+})=0.32\pm0.02 \text{ eV}. \quad (3)$$

All of the experiments from which these values were inferred used temperatures T below 250 K. Both deep-level transient spectroscopy (DLTS) experiments and electron paramagnetic resonance (EPR) experiments were employed to deduce these values. Some of these data are shown in Fig. 1, which has been adapted from Ref. 1. It is most likely that the values of ΔH_m for V^- and for V^+ are intermediate between those for V^{2-} and V^0 and for V^0 and V^{2+}, respectively.

Although there should be no doubt of the validity of the values of ΔH_m just given for low temperatures, the author has for several years argued that they do not apply at much higher temperatures.[2-6]

The crux of the author's previously published arguments is that at high T $\Delta H_m(V)$ is dominated by the kinetic energy of the atom that actually moves when we say the vacancy hops; potential energy may dominate at low T. Indeed, in Ref. 2, the approximate magnitude of $\Delta H_m(V)\sim0.36$ eV was derived from consideration of the minimal distortion energy required to move a Si atom from one site to another around a single vacancy cavity. However, such values are too small to allow the Si atom to make its transit in one period of a zone-boundary phonon. The kinetic energy alone required for a Si atom to hop the nearest-neighbor distance in a zone-boundary phonon period is

$$H_k(T>\Theta_D)=1.18 \text{ eV}, \quad (4)$$

where $\Theta_D=670$ K for Si is the Debye temperature. The zone-boundary period is the characteristic time for the fluctuations of the instantaneous potential in any particular unit cell of the specimen when T is great enough that such phonons are excited with high probability. This occurs roughly when T is greater than Θ_D.

As T is reduced below Θ_D, the characteristic time for thermal fluctuations of the potential slows until, at $T=0$ K, it ceases altogether. It may approximately be scaled as $(\Theta_D)/T$. Thus, we might estimate, that for normal

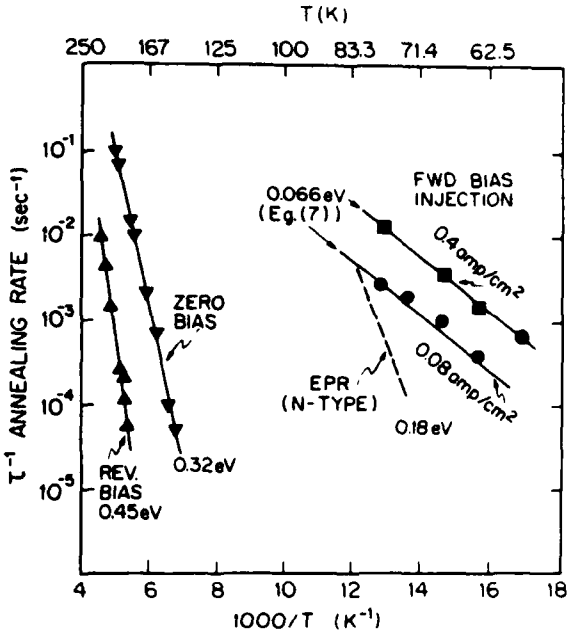


FIG. 1. Annealing kinetics for vacancy in Si determined by DLTS and EPR studies reviewed by Watkins (Ref. 1). The 0.18-eV Arrhenius is from EPR studies of n-type Si. All other data is for p-type Si. The two Arrhenius lines for forward bias are according to Eq. (7).

thermal conditions (i.e., no injection of excess carriers by light or currents), the minimum kinetic energy that a Si atom must have at the onset of a successful hop to a nearest-neighbor vacancy site is

$$H_k(T<\Theta_D)=(T/\Theta_D)^2\times 1.18\ \text{eV} . \quad (5)$$

For $T=250$ K, Eq. (5) gives $H_k=0.164$ eV. As this is less than the empirical value for ΔH_m obtained for all of the charge states of V when $T<250$ K, one might conclude that the activation enthalpy is dominated by potential energies in this regime. Evidently, this is so. It also seems that the potential energy contributions must be equal to or less than these empirical values. But then these potential energies are negligible compared with H_k when $T\gg\Theta_D$, as is the case for most atomic diffusion and device-processing experiments.

The physical picture is as follows. At high T the thermal vibrations of atoms surrounding a vacancy, or any other point defect where atomic migration can occur, fluctuates with a characteristic period given approximately by the bulk zone-boundary-phonon period. Periodically these thermal fluctuations of the surrounding atoms conspire to provide a path for a particular atom to hop from its initial site to that occupied by the vacancy at the cost of little potential energy. However, the venue for this circumstance is quite limited (to the zone-boundary-phonon period). Thus, the atom that is to make the jump must have sufficient kinetic energy to finish the jump before this easy path is closed off by the continuing thermal motion of the surrounding atoms. On the other hand, at low T when the zone-boundary phonons are not excited, the jumping atom may take much more time and the measured activation barrier may be dominated by the potential energy. The identification of $\Delta H_m(V)$ with H_k is called the "ballistic model" of vacancy migration.

The reader must realize that the ballistic model is a massive simplification of what is inherently a very complicated many-body phenomenon. In reality all the atoms of the solid are in motion and the various migration events that occur must involve some motion of the surrounding atoms, as well as the flight of the hopping atom from its initial to its final position. As such, the only justification for the ballistic model can be empirical. As has been discussed particularly in Refs. 5 and 6, with only one empirically adjusted parameter, that was fixed in 1975 (Ref. 3) to data for Ge at a value identical to that used for fcc structure noble-gas solids, the ballistic model has been found to work much better than one would initially expect in several semiconductor applications and for O vacancy migration in MgO. The reader is referred to Refs. 3–6 for the details of that justification. Here we briefly note several facts that can be cited as evidence for the assumption that kinetic energies dominate vacancy hopping activation energies at high temperatures. In the first place, in Ref. 3 it was shown that the simple kinetic-energy expression,

$$H_k=Mv^2/2 . \quad (6)$$

where M is the mass of the mass of the atom that jumps and v is its velocity assuming it moves from its initial lattice site to its final lattice site in a zone-boundary-phonon period, gives good agreement with the available experimental values of vacancy migration enthalpies in almost all cases of elemental crystals (metals and insulators as well as semiconductors) where values have been determined (at high temperatures). The values for Si and Ge obtained from (6) are 1.18 and 1.11 eV, respectively, which are in agreement with high-temperature values obtained for those crystals. The case is particularly clear in the instance of Ge where host-isotope-diffusion experiments have been done[7,8] to determine the isotopic mass dependence of the rate of diffusion. These experiments show $\Delta H_m(V_{Ge})$ to be about 85% kinetic energy in character. (There is in Ge a discrepancy between vacancy migration activation energy at low T and at high T that is very similar to the discrepancy for Si, but in the case of Ge the high-T value of about 1.1 eV is much more clearly established. See Refs. 2 and 3 and references therein.) The analogies between Si and Ge are so close and the chemical trends in material parameters are so clear that it is difficult for this author to believe that vacancy behavior could be radically different in the two cases. The assumption of the kinetic energy dependence of vacancy hopping energies has also been tested in group III-V compounds such as InP, where it gives predicted values for In-vacancy hopping and for P-vacancy hopping that differ from each other by a factor of 4 and are both in quantitative agreement with observed values.[5,6] [It should be noted that values consistent with the analog of Eq. (4), not of Eq. (5), have been found also for experiments[5,6] with $T<\Theta_D$ in InP. We return to this point later.] Moreover, analysis of the distribution of E centers, V-P_{Si} pairs, in Si produced by rapid quenching from temperatures close to (but below) the melting point have been interpreted[9] to conclude an activation energy for single vacancy migration of order 1.2 eV.

Consideration of the kinetic energy also provides a simple explanation for the dominance[5] of vacancy–first-nearest-neighbor hopping over second-nearest-neighbor hopping in III-V compounds, despite the cost of forming the antisite defects. The formation of antisite defects would be avoided if the vacancy were to hop to a second-nearest-neighbor site, i.e., one on its own sublattice. However, $\frac{8}{3}$ more kinetic energy would be required to accomplish this (in the same time) because the second-nearest-neighbor distance is a factor of $(\frac{8}{3})^{1/2}$ greater than the nearest-neighbor distance. Thus, for Si it would require 3.2 eV to hop to a second-nearest-neighbor site if it costs 1.2 to hop to a first-nearest-neighbor site.

Returning now to low-T phenomena in Si and to Fig. 1, we see that when a p-type Si DLTS sample is put into forward bias, so that electrons are injected into the conduction band (e_c^-), there is a very great enhancement of the vacancy annealing rate. Evidently the rate of vacancy migration is enhanced by the same large factor. The effect is called recombination-enhanced vacancy migration (REVM).

It is often stated in the literature regarding REVM that the process is independent of T and therefore athermal. However, we see in Fig. 1 a clear variation in the REVM rate with T; for a constant injection current $j=0.4$ A/cm^2 it varies by a factor of approximately 20 as T

varies from 60 to 80 K. Indeed, the lines in Fig. 1 associated with the forward bias data are the simple Arrhenius expression

$$R(T,j,)\propto j\exp(-\Delta H_{\mathrm{REVM}}/kT)\ , \tag{7}$$

where $\Delta H_{\mathrm{REVM}}=0.066$ eV. Due to inexactitude of the method, this value should probably be quoted as 0.07 eV. We see that only one datum, that for the lowest T and lower j, deviates appreciably from Eq. (7). It should be kept in mind that this result was obtained in p-type Si and is almost certainly dependent upon the charge state of the vacancies. It is not clear whether Eq. (7) pertains to recombination at V^{2+} or at V^{+}; the former is the more prevalent in thermal equilibrium, particularly because of the negative-U character[1] of the donor levels of vacancies in Si, but the injection of electrons in the experiment produces a nonequilibrium population of V^{+}, which can also trap injected electrons and produce REVM. The conversion of V^{+} to the more stable V^{2+} is rather slow.[1]

REVM is also observed in DLTS experiments with n-type Si and, until recently, this was accepted as evidence that there was an effective REVM mechanism for V^{2-}, or for V^{-}. However, as described in Ref. 1, there is now convincing evidence that what REVM does occur in n-type samples is occurring at positively charged vacancies, V^{2+} or V^{+}, which are created as a consequence of the injection of holes (h^{+}'s), and not at the V^{2-}'s or V^{-}'s at all. It seems possible that REVM might also be occurring at neutral vacancies, but there is no firm evidence of this.

In fairness to those who have called the Si REVM effect[1] athermal, it should be noted that there is in fact a detectable rate of REVM even at $T=4$ K where Eq. (7) implies $R(T,j)$ should be effectively zero. It seems that the net REVM effect must have two distinct components.

Discussions of REVM distinguish two classes of mechanism. (1) A "kick" mechanism[10–13] in which the energy of recombination of an electron e^{-} and a hole h^{+} at the defect H_R site dumps substantial energy into local modes of the lattice at the defect site in some manner that provides either part or all of the enthalpy required for the defect to migrate. (2) A charge-state fluctuation or "Bourgoin" mechanism[12] in which the defect has different configurations in two (or more) charge states and the passage of the current j (with or without recombination) causes transitions among these. If the stable configuration for one charge state is at or near the saddle point for migration of another, fluctuations of the charge state may induce the migration.

The Bourgoin charge-state fluctuation mechanism is usually said to be athermal but might be thermal (with an activation enthalpy less than that without injection) if some thermal energy must be combined with the fluctuation in equilibrium configuration to achieve the migration event. Problems associated with attempts to explain the data for vacancies in Si with the Bourgoin mechanism are discussed in Ref. 1 in terms of the Jahn-Teller configurations of the vacancy in its various charge states, which have been determined by EPR. The problems seem to be severe.

The "kick" mechanism will be thermal provided that H_R, or that fraction $F\le 1$ of it that is delivered to the mode that is effective in moving the vacancy, is less than that required for the interesting event ΔH_m in this case. Thus,

$$\Delta H_{\mathrm{REVM}}=\Delta H_m-FH_R\ , \tag{8}$$

because an extra amount of thermal enthalpy H_{REVM} must be obtained from the background phonon bath. When the "kick" mechanism occurs and a finite ΔH_{REVM} is measured in compound semiconductors, it is often found[10,11] that the normal thermal-activation-enthalpy barrier is reduced by an amount essentially equal to H_R. Of course, this implies that essentially all of H_R went into the mode that knocks the vacancy over to the next site. In Si, Watkins *et al.* have shown[13] interstitial Al double donors, which was created by electron irradiation, exhibit recombination-enhanced migration with an F factor greater than 90%. Thus,

$$F\simeq 1\ , \tag{9}$$

is usually found, although it is not yet clear why.

Let us now consider how large H_R is. The band gap enthalpy $\Delta H_{cv}=1.170$ eV for Si at $T=0$. It increases[14] slowly with T; at 70 K, the midpoint of the REVM Arrhenius in Fig. 1,

$$\Delta H_{cv}(T=70\ \mathrm{K})=1.173\ \mathrm{eV}\ . \tag{10a}$$

At the melting point of Si, 1685 K,

$$\Delta H_{cv}(T=1685\ \mathrm{K})=1.329\ \mathrm{eV}\ . \tag{10b}$$

H_R is generally less than ΔH_{cv} because true band-to-band recombination is extremely rare and is unlikely to affect vacancy hopping when it does occur. Two modes of recombination that might affect kick REVM can be distinguished. (1) The e^{-} and h^{+} may form an exciton and the exciton may annihilate itself at a vacancy which itself might be in any of its five charge states because the exciton itself is neutral. The annihilation would be facilitated by and well coupled to the vacancy because the localized electronic states of the V break the symmetry of the Si lattice. (As Si is an indirect-gap semiconductor, band-to-band recombination cannot occur in the perfect lattice without phonon emission or absorption to conserve crystal momentum.) The binding energy of the exciton at the vacancy is probably quite close to that in the bulk crystal, which is[14] 14.7 meV. Thus, for the exciton mechanism we have for exciton annihilation

$$H_R(\mathrm{exc},T=70\ \mathrm{K})=1.173-0.015\ \mathrm{eV}=1.158\ \mathrm{eV}\ . \tag{11}$$

(2) The other mode is ion annihilation. Either e^{-} in the conduction band, e_c^{-}, or h^{+} in the valence band h_v^{+} can annihilate one of the corresponding charges at an ionized vacancy. The ionized vacancy may be regarded as having trapped one or two e^{-}'s or h^{+}'s with a binding energy equal to the depth of the corresponding ionization level E_t away from the appropriate band edge. For V^{-} and V^{2-} the appropriate band edge is that of the conduction band, which we denote H_t and recall it is the enthalpy

gap[4,14] and not the more commonly discussed (free energy) band gap ΔE_{cv} that is relevant when calculating the slope of an Arrhenius plot. (However, with $T < 100$ K, the difference between enthalpies and free energies is not large on the scale of experimental and theoretical uncertainties. The difference will be significant when we consider high-T effects.) For V^+ and V^{2+} the appropriate band edge is that of the valence band, which we denote H_v. Thus, the enthalpy released when a h_v^+ annihilates a trapped electron at a V^{i-}, with $i=1$ or 2, is

$$H_R(\text{ion}, V^{i-}, T) = \Delta H_{cv}(T) - [H_c(T) - E_i(V^{i-}, T)] \quad (12)$$

and correspondingly when an e_c^- annihilates a trapped hole at a V^{i+}, again with $i=1$ or 2,

$$H_R(\text{ion}, V^{i+}, T) = \Delta H_{cv}(T) - [E_i(V^{i+}, T) - H_v(T)] \, . \quad (13)$$

Van Vechten and Thurmond have discussed[4,15] the problem of the variation of the vacancy ionization levels relative to the band edges. For $T < 100$ K this variation is insignificant. The following values are taken literally from Watkins' ionization level diagram:

$$E_i(V^+) - H_v = 0.05 \text{ eV}, \quad H_c - E_i(V^-) = 0.65 \text{ eV} \, ,$$
$$E_i(V^{2+}) - H_v = 0.13 \text{ eV}, \quad H_c - E_i(V^{2-}) = 0.29 \text{ eV} \, . \quad (14)$$

However, Watkins warns that the two acceptor levels have not been determined with precision; one can be sure that they are both more than 0.17 eV below E_c, but beyond this there is uncertainty. Despite this caveat, we here take these literal values in order to illustrate clearly the points of the discussion. The qualitative points of the low-T discussion are not affected by the precise values of these acceptor levels, so long as they are deeper than 0.17 eV. The quantitative consequences of the level positions will be seen when we discuss the high-T vacancy migration enthalpy. We will find no conflict with experiment in the high-T data when we use the literal values. These same values were used successfully to account for positron annihilation lifetimes in thermally generated vacancies in Ref. 16.

Accepting these values, we calculate from (12) and (13) that

$$H_R(\text{ion}, V^+, T = 70 \text{ K}) = 1.123 \text{ eV} \, , \quad (15a)$$

$$H_R(\text{ion}, V^{2+}, T = 70 \text{ K}) = 1.043 \text{ eV} \, , \quad (15b)$$

$$H_R(\text{ion}, V^-, T = 70 \text{ K}) = 0.523 \text{ eV} \, , \quad (15c)$$

$$H_R(\text{ion}, V^{2-}, T = 70 + \text{K}) = 0.883 \text{ eV} \, . \quad (15d)$$

Obviously, there is an uncertainty attached to each of these values which probably exceeds 0.02 eV. (The last digit is shown only for clarity to remind the reader of the increase in ΔH_{cv} with increasing T.)

If one assumes that the normal thermal values of the activation enthalpies at $T = 70$ K, the ΔH_m's is given in Eqs. (1)–(3), are relevant to REVM around $T = 70$ K, then one is faced with a rather mysterious situation. All the various H_R's are much larger than any of the ΔH_m's for the various charge states of the vacancy. Thus, with the common experience that $F \simeq 1$ in Eq. (8), we would expect REVM to occur for all charge states of vacancy and to be athermal, i.e., independent of T, because the enthalpy being dumped into the local models at the vacancy is much more than required to produce the hop. To account for the absence[1] of REVM for V^- and V^{2-}, we must have $F(V^-) < 0.34$ and $F(V^{2-}) < 0.20$ for the ion modes and $F(V^{i-}) < 0.16$ for the exciton mode; these are uncharacteristically small values. To account for the presence of REVM with $\Delta H_{\text{REVM}} = 0.07$ eV at either V^+ or V^{2+}, we have to have for the ion modes $F(V^+) = 0.226$, so that $FH_R = \Delta H_m - 0.07$ eV, or $F(V^{2+}) = 0.240$. If the 0.07-eV activation enthalpy pertains to the exciton mechanism, we need $F = 0.216$ for a positively charged vacancy or $F(V^0) = 0.328$ for the neutral vacancy. All these values are uncharacteristically small and mysterious. We would further have to account for the truly athermal REVM process, which requires a slightly larger value of F for one of the modes at either V^0, V^+, or V^{2+}. No explanation of any of this is given in Ref. 1 nor, to this author's knowledge, anywhere else.

However, suppose instead that we assume the high-T ballistic model value $\Delta H_m(V) = 1.18$ eV, from Eq. (4) is at least approximately the value relevant to the REVM effect for $T = 70$ K. This assumption is justified by the considerations that (1) the recombination event occurs in a time short compared with the zone-boundary-phonon period and the H_R released from it is rapidly dissipated in a similar time; and (2) the zone-boundary phonons must certainly be excited by the recombination if an atomic hop is to be produced. Let us also assume that for each one of the possible modes and charge states,

$$F = 1.0 \, . \quad (16)$$

Furthermore, although the high-T ballistic-model value of ΔH_m was never fixed with great precision and although the attribution of that value to REVM at 70 K must certainly be regarded as approximate, let us for the moment assume the relevant value of ΔH_m is exactly 1.180 eV. Then we calculate that for the exciton mode at any charge state of the vacancy

$$\Delta H_{\text{REVM}}(\text{exc}) = 1.180 - 1.158 \text{ eV} = 0.022 \text{ eV} \quad (17)$$

and for the ion annihilation mode we calculate

$$\Delta H_{\text{REVM}}(\text{ion}, V^+, T = 70 \text{ K}) = 1.180 - 1.123 \text{ eV} = 0.057 \text{ eV} \, ; \quad (18a)$$

$$\Delta H_{\text{REVM}}(\text{ion}, V^{2+}, T = 70 \text{ K}) = 1.180 - 1.043 \text{ eV} = 0.137 \text{ eV} \, ; \quad (18b)$$

$$\Delta H_{\text{REVM}}(\text{ion}, V^-, T = 70 \text{ K}) = 1.180 - 0.883 \text{ eV} = 0.657 \text{ eV} \, ; \quad (18c)$$

$$\Delta H_{\text{REVM}}(\text{ion}, V^{2-}, T = 70 \text{ K}) = 1.180 - 0.523 \text{ eV} = 0.297 \text{ eV} \, . \quad (18d)$$

We immediately see that if we reduce our previous esti-

mate of the relevant value of ΔH_m by only as much as 2%, from 1.180 eV to 1.158 eV, Eq. (17) becomes $\Delta H_{\text{REVM}}(\text{exc})=0$ for whichever charge state of vacancy may be the site of the exciton annihilation mode of REVM. This is just what is required to account for the truly athermal component of REVM that occurs at either V^0, V^+, or V^{2+}. (This author favors the assumption that the V^0 is the site of the athermal REVM.) Such a 2% adjustment is obviously well within the uncertainty of the hypothesis. If the downward adjustment is any more than 2%, then we still have an explanation for the athermal component of REVM. The only remaining point is the apparent absence of this mode at V^- and V^{2-}.

If the assumed relevant value for ΔH_m is to be reduced slightly from 1.180 eV in Eq. (17), then the same adjustment should be made in Eqs. (18). We could bring the calculated value of $\Delta H_{\text{REVM}}(\text{ion}, V^{2+})$ into agreement with the observed value $\Delta H_{\text{REVM}}=0.07$ eV by making the adjustment 0.067 eV or 5.7% instead of 2%. This would also make $\Delta H_{\text{REVM}}(\text{ion}, V^+)<0$, i.e., give us an alternative or an additional explanation for the athermal component of REVM, provided we retain Eq. (16). It is easy to believe the 1.180 eV estimate is uncertain to 5.7% also. However, the empirical value of ΔH_{REVM} and perhaps also the deviation of the lowest-T datum from the Arrhenius in Fig. 1 might also be explained by the effect of the combination of ion annihilation at V^+ and at V^{2+}; the one gives a contribution with a slightly smaller enthalpy while the other gives a contribution with a slightly larger enthalpy.

Now consider what our calculation implies for V^- and for V^{2-}. We see that the activation enthalpies for the ion annihilation mode of REVM at V^{2-} and at V^- are larger than the activation enthalpy for normal thermal migration (at least for V^{2-} since the value for V^- is not well established). A little consideration shows that if the normal thermal process, which is always occurring in competition to any recombination-enhanced process, has a smaller activation enthalpy it will almost certainly swamp out the recombination-enhanced process. The normal thermal-migration event can occur at any moment at any of the V^{2-}'s while the REVM event can only occur when and where there is recombination at V^{2-}. Thus, the preexponential factor, or attempt frequency, should be greater for the thermal process. When the thermal process has a larger preexponential factor as well as a much larger exponential factor (due to the lower activation enthalpy), the product must be much larger than that for the REVM process. As noted above REVM indeed is not observed at V^{2-}'s. The same conclusion follows for V^-.

Recall that we noted above that for InP values for the activation enthalpy for In and P vacancy diffusion were measured[5,6] that were in close agreement with those calculated with the ballistic model for high T and that these same values were also found even when $T \ll \Theta_D$. One might think that observation would cast doubt upon Eq. (5) if it were not for the fact the data used for those low-T studies involved electrical or optical injection. Thus, instead the observation is further evidence for the validity of the assumption that it is the high-T–kinetic-energy value for ΔH_m that should be used to estimate the total enthalpy required for the REVM activation enthalpy.

Finally let us consider what implications REVM may have on the behavior of vacancies at high T. The issue arises because minority carriers are generated thermally and will recombine at vacancies even in the absence of electrical or optical injection and because positron annihilation experiments[17] firmly demonstrate that vacancies with static properties very similar to those observed at low temperature are present for $T > 1200$ K in thermal equilibrium in Si at significant concentrations (greater than 10^{16} cm^{-3}). To appreciate that REVM should be a significant effect under high-T thermal conditions, we will now make two points via simple estimates: (1) The minority-carrier density, and thus the thermal rate of recombination, is much higher for such thermal experiments than for the low-T REVM experiments;[1-13] and (2) the "attempt frequency" for thermal REVM is at least comparable to that for the ordinary thermal process in which the enthalpy is obtained from the phonon bath with no assistance from the free carriers.

(1) The intrinsic carrier concentration in Si at its melting point, 1685 K, is about[14] 2×10^{19} cm^{-3}. (This value is typical of most tetrahedral semiconductors at their melting points.) This value is large enough to overwhelm the doping in all but the most heavily doped samples studied at high temperatures. Thus, the minority carrier concentration may be estimated at the intrinsic carrier concentration for most high-T studies. On the other hand, a forward bias current of 0.4 A/cm^2 into a sample with a minority-carrier diffusion length of 5 μm and a minority-carrier lifetime of 1 μsec implies a minority-carrier density in the junction region of order 5×10^{15} cm^{-3}.

(2) To estimate the "attempt frequency" for REVM, we may first note that the thermal velocity of a carrier at $T=1685$ K is 2.8×10^7 cm/sec and the Coulomb-capture cross section[18] of a singly charged vacancy is 8.7×10^{-14} cm^2 so the rate at which one of the 2×10^{19} cm^{-3} oppositely charged carriers impinge on any one of the singly charged vacancies is, according to the Coulomb cross-section criterion, 4.8×10^{13} sec^{-1}. This may be compared with the normal thermal attempt frequency which is usually taken to be $k\Theta_D/h=1.4\times10^{13}$ sec^{-1}. As the Coulomb-capture cross section for a doubly charged vacancy is four times that for a singly charged vacancy, the thermal REVM attempt frequent at such vacancies will be correspondingly greater. However, the Coulomb cross-section criterion is probably something of an overestimate under these conditions. There are so many carriers of both signs that a single vacancy may interact with more than one simultaneously and the time the carrier spends falling into the Coulomb well may not be negligible compared with the time it spends thermally diffusing before it reaches the point that the attractive Coulomb potential of the charged vacancy exceeds the $\frac{1}{2}kT$ radial thermal average value. To quantify this caveat we note that the radius of the coulomb cross section is 1.66 nm so the corresponding volume is 1.93×10^{-20} cm^3 and the random probability of a carrier being in that volume is about 10% for either type. We

conclude that the attempt frequency for thermal REVM is at least comparable to, if not larger than, the ordinary thermal phonon attempt frequency.

Given that thermal REVM, TREVM, is occurring with a frequency comparable to ordinary thermal-phonon attempt frequency, we must conclude that TREVM makes a significant contribution to the migration of vacancies at processing temperatures which becomes dominant as T approaches 1685 K, the melting point. This follows because the probability of a hop actually occurring at each and any attempt is much larger for TREVM, due to the kick gained from the recombination, than for thermal-phonon-activated vacancy migration TPAVM. If one accepts that this author is correct in that the total enthalpy required for vacancy migration at $T>670$ K is about 1.18 eV [Eq. (4)] and not the values measured with $T<250$ K [Eqs. (1) to (3)], then at 1685 K the TRAVM hop occurs only for one hop in a number of order $\exp(1.18\ \text{eV}/1685k)=3.4\times10^3$ attempts. The uncertainty is due to the entropy-of-migration factor, which cannot be larger[2,3] than about 10 (not the entropy-of-vacancy-formation factor). By contrast, we see that, due to the increase in ΔH_{cv} with T, the exciton provides enough kick to boost the vacancy through a 1.18-eV hop with no need for added enthalpy from the thermal bath. That is, from Eq. (10b) we have

$$H_R(\text{exc}, T-1685)=1.329-0.015\ \text{eV}=1.314\ \text{eV} \quad (19)$$

so that

$$\Delta H_{\text{TREVM}}(\text{exc})<0\ . \quad (20)$$

Furthermore, according to the estimate[15] of Van Vechten and Thurmond, the variation of the vacancy ionization levels is such that H_R is increased for each of these by the same amount, 0.159 eV, as is the ΔH_{cv}, and is H_R for the exciton. Thus, all the ΔH_{TREVM} for ion annihilation modes of all the charge states of the vacancy are also 0.159 eV less at $T=1685$ K than the corresponding values of ΔH_{REVM} at $T=70$ K, Eqs. (18a) and (18b). This implies $\Delta H_{\text{TREVM}}<0$ for the ion annihilation mode at V^+ and at V^{2+}; we thus expect a factor of order 3.4×10^3 increase (over the extrapolation of the ordinary TPAVM rate from Ts low enough that TREVM is not important due to the scarcity of minority carriers in the absence of injection) in the rate of V^{2+} and/or V^+ migration and in the rate of V^0 migration if indeed the exciton-annihilation mode is effective at V^0s.

In addition, we also expect to see at high Ts a TREVM effect on V^- and V^{2-}, despite the evident absence of a REVM effect at low T, as noted above and in Ref. 1. This is because the competing strictly thermal process TPAVM is (at least by this author) assumed to have a total activation enthalpy, 1.18 eV, that is so large that the thermal process does not swamp the recombination enhanced process (as we have argued occurs at low T). However, for the negative vacancies we have positive values for ΔH_{TREVM}

$$\Delta H_{\text{TREVM}}(\text{ion}, V^-, T=1685\ \text{K})=0.498\ \text{eV}\ , \quad (21)$$

$$\Delta H_{\text{TREVM}}(\text{ion}, V^{2-}, T=1685\ \text{K})=0.138\ \text{eV}\ . \quad (22)$$

Consider now the corresponding changes in the slopes of the Arrhenius plots of vacancy migration in the five charge states due to the TREVM effect. Any change in slope must be positive, i.e., correspond to a larger net activation enthalpy because the effect is to increase the net rate of migration to an increasing degree as T increases. Looking at this another way, we note that it must always cost enthalpy to create the e^- and the h^+ that can recombine to make the vacancy hop and we will never get all this enthalpy back in a reduction of the hopping activation enthalpy. With a little consideration one sees that indeed the slope increases by just $\Delta H_{cv}-FH_R$, i.e., the amount of the cost of the e^--h^+ pair that was not regained in the mode that boosts the vacancy in its hop. As we are assuming $F=1$ to a good approximation (as is usually found), the increase in slope for the exciton-annihilation mode is only the exciton binding energy about 0.015 eV, which would be hard to detect. For the ion modes at V^+ the increase in slope would be only $E_i(V^+)-H_v$, about 0.05 eV, which also could probably not be detected. At V^{2+} the corresponding increase would be about 0.13 eV. [Recall Eq. (14).] However, at V^- and at V^{2-} the slope would increase by 0.65 and by 0.29 eV, respectively, due to the ion mode. Note that if we are to assume somewhat different values for the accepter ionization levels at Eq. (14), these values would change by the same amounts. As was noted above, the exciton annihilation does not seem to occur at V^- and V^{2-}, so the net slope for V^- and V^{2-} migration should increase by these amounts.

Now it is well established that the rate of self-diffusion in Si does not obey[19] a simple Arrhenius. Around $T=1200$ K the slope for self-diffusion (which contains the enthalpy of formation of the defect mediating the self-diffusion as well as the enthalpy of migration of that defect) is (according to Demond *et al.*, who seem to have the very best experiment[19] in the field) 4.15 ± 0.25 eV, while around 1600 K the slope is about 5.0 eV. There is a corresponding increase in the preexponential factor for self-diffusion from a value between 0.3 and 30 cm^2/sec at 1200 K to about $2\times10^3\ \text{cm}^2/\text{sec}$ at 1600 K. Particularly in view of the fact that positron annihilation experiments show large concentrations of single vacancies to be present in thermal equilibrium at these Ts, there is also firm evidence that at least a substantial fraction of the self-diffusion in Si is mediated by single vacancies. (Please note that recent studies of the growth of stacking faults[20,21] favor a dominance of vacancy mechanisms over any self-interstitial mechanism.) Furthermore, the most prevalent[16] charge state of vacancy in lightly doped Si with $1200\ \text{K}<T<1685$ K, is V^-. Thus, a large fraction, if not all, the increase in the slope of the Si self-diffusion Arrhenius plot can be attributed to the 0.65 eV increase that occurs with the ion annihilation mode TREVM effect at V^-'s, provided we take Watkins' ionization-level diagram literally. Moreover, a large fraction, if not all, of the increase in the preexponential factor for Si self-

diffusion can be attributed to the TREVM effect at vacancies in all their charge states. Of course, similar thermal recombination enhancements are expected for divacancies,[21] and for any other point defect that may contribute significantly to Si self-diffusion, provided they also have a recombination-enhanced migration mechanism.

This work was supported in part by the Air Force Office of Scientific Research under Contract No. AFOSR 86-0309.

[1]The latest comprehensive review is in G. D. Watkins, in *Deep Centers in Semiconductors*, edited by S. T. Pantelides (Gordon and Breach, New York, 1986), Chap. 3, p. 147.
[2]J. A. Van Vechten, Phys. Rev. B **10**, 1482 (1974).
[3]J. A. Van Vechten, Phys. Rev. B **12**, 1247 (1975).
[4]J. A. Van Vechten, in *Handbook on Semiconductors*, edited by S. P. Keller (North-Holland, Amsterdam, 1980), Vol. 3, Chap. 1.
[5]J. A. Van Vechten and J. F. Wager, Phys. Rev. B **32**, 5259 (1985).
[6]M. T. Juang, J. F. Wager, and J. A. Van Vechten, J. Electrochem. Soc. **135**, 2019 (1988); **135**, 2023 (1988).
[7]D. R. Campbell, Phys. Rev. B **12**, 2318 (1975).
[8]D. Shaw, Phys. Status Solidi B **72**, 11 (1975).
[9]A. Chantre, M. Kechouane, and D. Bois, Physica **116B**, 547 (1983).
[10]J. D. Weeks, J. C. Tully, and L. C. Kimerling, Phys. Rev. B **12**, 3286 (1975).
[11]A. M. Stoneham and R. H. Bartram, Solid State Electron. **21**, 1325 (1978).
[12]J. C. Bourgoin and J. W. Corbett, Radiat. Eff. **36**, 157 (1978).
[13]G. D. Watkins, A. P. Chatterjee, R. D. Harris, and J. R. Troxell, Semicond. Insul. **5**, 321 (1983).
[14]C. D. Thurmond, J. Electrochem. Soc. **122**, 1133 (1975).
[15]J. A. Van Vechten and C. D. Thurmond, Phys. Rev. B **14**, 3539 (1976).
[16]J. A. Van Vechten, Phys. Rev. **33**, 2674 (1986).
[17]S. Dannefaer, P. Mascher, and D. Kerr, Phys. Rev. Lett. **56**, 2195 (1986).
[18]M. Lax, Phys. Rev. **119**, 1502 (1960).
[19]F. J. Demond, S. Kalbitzer, H. Mannsprenger, and H. Damjantschitsch, Phys. Lett. **93A**, 503 (1983).
[20]J.-I. Chikawa, in *Defects and Properties of Semiconductors: Defect Engineering*, edited by J. Chikawa, K. Sumino, and K. Wada (KTK Scientific, Tokyo, 1987), p. 143.
[21]K. Wada and N. Inoue, in *Defects and Properties of Semiconductors: Defect Engineering*, Ref. 20, p. 169.

Reprinted from Journal of the Electrochemical Society
Vol. 135, No. 8, August 1988
Printed in U.S.A.

Phosphorous Vacancy Nearest Neighbor Hopping Induced Instabilities in InP Capacitors

II. Computer Simulation

M. T. Juang, J. F. Wager, and J. A. Van Vechten*

Department of Electrical and Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, Oregon 97331

ABSTRACT

Drain current drift in InP metal insulator semiconductor devices display distinct activation energies and pre-exponential factors. We have given evidence that these result from two physical mechanisms: thermionic tunneling of electrons into native oxide traps and phosphorous vacancy nearest neighbor hopping (PVNNH). We here present a computer simulation of the effect of the PVNNH mechanism on flatband voltage shift *vs.* bias stress time measurements. The simulation is based on an analysis of the kinetics of the PVNNH defect reaction sequence in which the electron concentration in the channel is related to the applied bias by a solution of the Poisson equation. The simulation demonstrates quantitatively that the temperature dependence of the flatband shift is associated with PVNNH for temperatures above room temperature.

It has been argued (1-6) that in III-V (but not in II-VI) compound semiconductors the energy of antisite defect formation is so small that the dominant mode of vacancy diffusion is via nearest neighbor hopping (nnh). Superlattice interdiffusion studies lend support to this assertion (7, 8). Further support for this claim was presented in the preceding paper (9) in which the high temperature activation energy associated with drain current drift (DCD) instabilities was determined experimentally to be 1.1-1.2 eV, in good agreement with the prediction (5) for phosphorous vacancy nearest neighbor hopping (PVNNH).

It is the purpose of this paper to describe the results of a computer simulation of the flatband voltage shift *vs.* the bias stress time for the PVNNH mechanism. The simulation is based on an analysis of the kinetics of the PVNNH defect reaction sequence. Electron trapping is accounted for in a semiempirical fashion, electron shielding is accounted for by a solution of the Poisson equation, and the kinetic rate equation is solved iteratively. The simulation agrees quite well with the experimentally determined temperature-dependence of the flatband voltage shift but deviates slightly from the experimental data at high temperature. Several reasons for this deviation are suggested.

Kinetics of PVNNH: A Computer Simulation

The central idea of PVNNH is that electrons in the channel of an InP metal-insulator-semiconductor (MIS) capacitor or metal insulator semiconductor field effect transistor (MISFET) are captured by the shallow acceptor levels which arise from an In atom nearest neighbor hop into a phosphorous vacancy, V_P

$$V_P^{+} + 4e = V_{In}\ In_P^{-2} \qquad [1]$$

This electron capture is manifest as DCD. It is assumed that P vacancies (because of their low enthalpy of formation) are present in the channel in concentrations (4, 5, 9) of order 10^{18} cm^{-3}. One of the four nearest neighbor In atoms surrounding the V_P may hop into the V_P, thus transforming the simple vacancy into a new defect complex consisting of an In vacancy and an antisite defect, as indicated in reaction [1]. Application of the law of mass action to this defect reaction results in

$$\frac{V_{In}\ In_P^{-2}}{[V_P^{+}]} \propto [e^{-}]^4 \qquad [2]$$

Thus, when an InP MIS capacitor or MISFET is biased into accumulation (*i.e.* the Fermi level is moved towards the conduction band, thus increasing $[e^{-}]$), Eq. [2] indicates that the equilibrium is shifted to the right-hand side of the defect reaction, Eq. [1]. Four of the accumulated electrons in the channel will be captured for each V_P annihilated. DCD is attributed to the loss of these electrons from the channel.

The PVNNH reaction may be broken up into a more detailed reaction sequence and the time constants at room temperature for the individual steps of the forward reaction are as follows [see Ref. (5) for details of how the time constants were estimated]

$$V_P^{+} + e = V_P^{o} \qquad \tau_a = 1\ \text{ps} \qquad [3a]$$

$$V_P^{o} = (V_{In}In_P)^{o} \qquad \tau_b = 0.5\text{s} \qquad [3b]$$

$$(V_{In}In_P)^{o} + e = (V_{In}In_P) \qquad \tau_c = 1\ \text{ns} \qquad [3c]$$

$$(V_{In}In_P) + e = (V_{In}In_P)^{-2} \qquad \tau_d = 3\ \mu\text{s} \qquad [3d]$$

$$(V_{In}In_P)^{-2} + e = (V_{In}In_P)^{-3} \qquad \tau_e = 8\ \text{ms} \qquad [3e]$$

The rate limiting step is [3b], in which the indium atom hops into the phosphorous vacant site.

As discussed below, this reaction sequence yields a rate equation which does not adequately describe the experi-

* Electrochemical Society Active Member.

mental flatband shift data. We will show, however, that a slightly different reaction sequence leads to an accurate simulation of the experimental data.

The kinematic equation for step [3b] is

$$\frac{d[(V_{In}In_P)^{n}]}{dt} = k_b[V_P^{o}] - k_b[(V_{In}In_P)^{n}] \quad [4]$$

where k_b and k_b are the rate constants for the forward and reverse reaction of step [3b] and where k_b is just the inverse of τ_b. Since the time constants of the other four steps are much smaller than the time constant of step [3b], we can treat step [3a] as if it were in equilibrium

$$K_a = \frac{[V_P^{n}]}{[V_P^{+}]\,[e^{-}]} \quad [5]$$

where K_a is the equilibrium constant for step [3a]. Once $(V_{In}In_P)^{n}$ is formed, it captures electrons very quickly and becomes $(V_{In}In_P)^{-3}$. In this case, Eq. [4] can be rewritten by replacing $[V_P]$ obtained in terms of $[V_P^{+}]$, $[e^{-}]$, and K_a from Eq. [5]. Similarly, $[(V_{In}In_P)^{n}]$ can be replaced by

$$[(V_{In}In_P)^{o}] = \frac{[(V_{In}In_P)^{-3}]}{K_cK_dK_e[e^{-}]^3} \quad [6]$$

and k_b can be calculated from $K_b = k_b/k_b$, where K_b, K_c, K_d, K_e are the equilibrium constants of steps [3b], [3c], [3d], and [3e], respectively. K_b can be estimated from the activation energy that is obtained from the experiment. K_c, K_d, K_e can be calculated from the energy that is released associated with the capture of electrons in these traps

$$-kT\ln K_a = -[E_f - E_t(V_P)] \quad [7]$$

$$-kT\ln K_b = 1.1\ \text{eV} \quad [8]$$

$$-kT\ln K_c = -[E_f - E_{t2}(In_P)] \quad [9]$$

$$-kT\ln K_d = -[E_f - E_{t1}(In_P)] \quad [10]$$

$$-kT\ln K_e = -[E_f - E_t(V_{In})] \quad [11]$$

The calculated K_b at 300 K and 350 K is 3.6×10^{19} and 1.5×10^{19}; the corresponding k_b is 5.6×10^{18} and 1.3×10^{19} for 300 and 350 K, respectively. The ionization levels of vacancies and antisite defects (5) in InP have been estimated by Temkin *et al.* (10) and Buisson *et al.* (11) to be 0.21, 0.17, 0.04, and 0.99 eV for V_{In}, E_{t1} (In_P), $E_{t2}(In_P)$, and $E_t(V_P)$, respectively. Using these values, with a doping concentration of 4×10^{15} cm^{-3}, the equilibrium constants estimated at 300 K are $K_c = 1.9 \times 10^{20}$, $K_d = 1.3 \times 10^{18}$, $K_e = 1.0 \times 10^{17}$, and at 350 K are $K_c = 4.8 \times 10^{16}$, $K_d = 6.5 \times 10^{14}$, $K_e = 1.7 \times 10^{14}$. According to these values, the reverse reaction is very small and the rate equation can be approximated by

$$\frac{d[(V_{In}In_P)^{-3}]}{dt} = K_aK_b[V_P^{+}][e^{-}] \quad [12]$$

However, if Eq. [12] is used in the simulation, with the procedure that is described later, the temperature dependence of the flatband shift cannot be fit to the experimental data. The amount of flatband shift obtained from the simulation at 350 K would be smaller by a factor of three compared to that obtained from the experiment.

Another possible reaction sequence is proposed to describe the kinetics of the reaction, as shown in Fig. 1. The solid line indicates the relative activation barrier that is associated with these individual reaction steps before the application of the gate bias. After the gate bias is applied, the quasi-Fermi level position is shifted toward the conduction band and the energy released that is associated with an electron captured by the trap is increased by an amount that is approximately the amount of shift in the Fermi level. Since the rate of electron capture by these traps is very fast, the barrier that an electron has to overcome is small. If the applied gate bias is slightly larger than

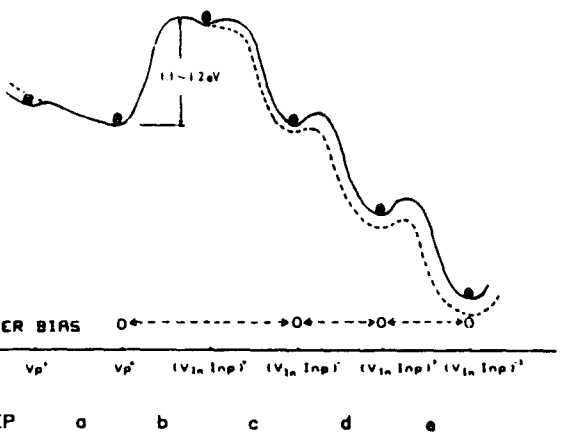


Fig. 1. Proposed relative reaction barriers for the PVNNH mechanism. Solid and dashed lines correspond, respectively, to before and after applying a gate bias. Only three stable configurations remain after the gate bias is applied.

the barrier for the forward reaction of step [3c], a new activation barrier curve representing this condition would be as indicated by the dashed line shown in Fig. 1. Note that the relative positions of the neutral species V_P^{o} and $(V_{In}In_P)^{o}$ are not influenced by the change in the Fermi level; the change in the Fermi level smears out the forward barriers for steps [3a] and [3c], which results in only three steps left in the reaction sequence

$$V_P^{o} + e^{-} = (V_{In}In_P)^{-} \quad [13]$$

$$(V_{In}In_P)^{-} + e^{-} = (V_{In}In_P)^{-2} \quad [14]$$

$$(V_{In}In_P)^{-2} + e^{-} = (V_{In}In_P)^{-3} \quad [15]$$

The rate limiting step is now [13] and the time constant for this forward reaction can be approximated by the time constant of step [3b]. Steps [14] and [15] are very fast compared to step [13]. Using a similar argument as for the five-step sequence, the reverse reaction rate is very small and can be neglected. The kinetic equation is then

$$\frac{d[(V_{In}In_P)^{-3}]}{dt} = k_c[V_P^{o}][e^{-}] \quad [16]$$

The number of trapped electrons associated with the formation of $(V_{In}In_P)^{-3}$ that is measured is only three instead of four because the first electron is captured even before the application of the accumulation bias because the Fermi level at zero bias is above the ionization energy of V_P. The equation for the trapped electron concentration is then

$$\frac{dN_{tr}}{dt} = 3k_c[V_P^{o}][e^{-}] \quad [17]$$

The electron concentration $[e^{-}]$ depends on the surface potential ψ_s. As trapping occurs, shielding due to the trapped electrons reduces the effective bias, and the surface potential decreases. $[e^{-}]$ is also a function of distance from the oxide-InP interface; it can be expressed in terms of the electric potential $\psi(x)$ in the InP. To find $\psi(x)$, the Poisson equation must be solved. If nondegenerate statistics are used, for the purpose of simplicity, it can be shown (12) that

$$E_s = \frac{2}{L_D}\left[(e^{\beta\psi_s} + \beta\psi_s - 1) + \frac{p_{no}}{n_{no}}(e^{-\beta\psi_s} + \beta\psi_s - 1)\right]^{1/2} \quad [18]$$

where E_s is the surface electric field. E_s can also be shown (13) to be

$$E_s = \frac{\epsilon_{ox}}{\epsilon_s}\left[\frac{V_g - \psi_s - \psi_{ms}}{X_n} - \frac{qN_e}{\epsilon_0\epsilon_{ox}}\left(1 + \frac{\bar{X}}{L_D}\right) - \frac{qN_e'}{\epsilon_0\epsilon_{ox}}\right] \quad [19]$$

where ψ_{ms} is the work function difference between the Al metal electrode and InP, $\beta = kT/q$, L_D is the Debye length, V_g is the gate voltage applied, X_o is the oxide thickness, N_e is the number of trapped electrons in the native oxide per square centimeter, $\bar{X}$ is the centroid of the trapped electrons from the interface

$$\bar{X} = \frac{\int_o^{L_D} XN_{tr}(x,t)dx}{N_{tr}(x,t)dx} \quad [20]$$

$$N_e = \int_o^{L_D} N_{tr}(x,t)dx \quad [21]$$

where $N_{tr}(x, t)$ is the concentration of the trapped electrons as a function of distance from the interface and as a function of time. The thickness of the accumulation layer was approximated as the Debye length. The electric potential can be approximated as

$$\psi(x) = xE_s/L_D \quad [22]$$

and electron concentration in the accumulation layer is then

$$[e^-] = N(x) = Nn \exp[\beta\psi(x)] \quad [23]$$

Equations [17]-[23] describe the trapping process. The flatband voltage shift, ΔV_{FB}, is

$$\Delta V_{FB} = \frac{qN_e}{C_o}\left(1 + \frac{\bar{X}}{L_D}\right) + \frac{qN_e'}{C_o} \quad [24]$$

where C_o is the oxide capacitance.

However, simulation of the flatband voltage shift still cannot be accomplished in a straightforward fashion. The contribution from electrons trapped in the native oxide, N_e', is estimated in an empirical fashion by extrapolating the low temperature experimental curves in order to deduce the amount of flatband shift at high temperature (*i.e.*, 300, 325, 350 K) which is due to native oxide trapping. This procedure amounts to assuming that the flatband shift is simply a superposition of the shift from the two mechanisms.

The simulation begins at zero time with the assumption that no electrons are initially trapped. E_s is solved to determine the electron concentration in the accumulation layer. The accumulation layer is divided into 50 sections in order to numerically accomplish the integrations indicated in Eq. [20] and [21]. It is assumed that N_{tr}, E_s, $[V_P^o]$, and $[e^-]$ are constant during the calculation time increment, Δt. N_{tr} is then

$$N_{tr}(t + \Delta t) - N_{tr}(t) = 3k, \{[V_P^o](x,t)\}\{[e^-](x,t)\}\Delta t/\tau \quad [25]$$

τ is the time constant for nearest neighbor hopping of In into a V_P and is given (5) by

$$1/\tau = \nu_o \exp(\Delta S_m/k) \exp(-\Delta H_m/kT) \quad [26]$$

where ν_o is the attempt frequency, ΔS_m is the activation entropy of the hop, ΔH_m is the activation enthalpy of the hop, and k is Boltzmann's constant. The phosphorous vacancy concentration is a function of distance into the accumulation layer because the consumption rate of the vacancy depends upon the electron concentration, which is also position dependent. $N_{tr}(x, t)$ is then integrated for the whole accumulation layer thickness to obtain the trapped electron density due to PVNNH, and then added to the amount calculated from the extrapolated value of electrons that are trapped in the oxide at the time of the calculation to obtain the total amount of trapped electrons. $\bar{X}$ and ΔV_{FB} are then calculated. For the next time increment, $\bar{X}$, N_e, and N_e' are substituted into Eq. [10] and a new surface electric field is calculated. The procedure is repeated until the desired simulation time period is achieved. A flow chart for the simulation is illustrated in Fig. 2.

The length of the calculation time increment, Δt, has to be chosen to be small enough such that E_s, $[V_P^o]$, and $[e^-]$ remain approximately constant, as assumed in the formulation. Time increments of 0.001, 0.1, and 1s were used for the three simulation temperatures, and yielded virtually the same results; deviations were only 0.5%.

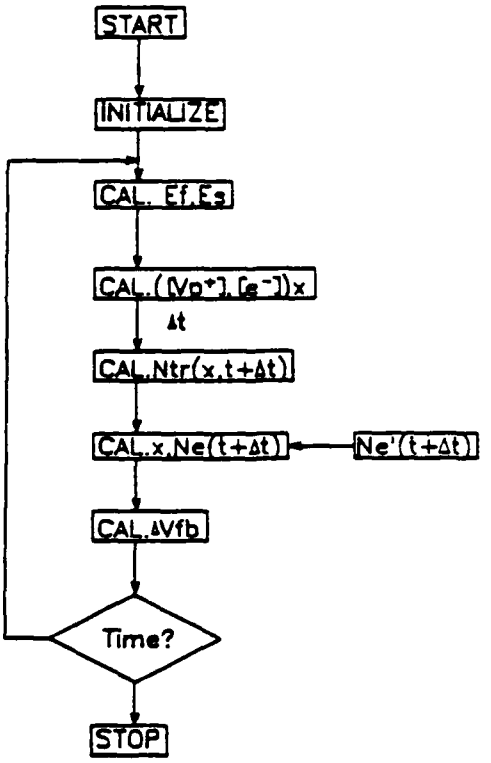


Fig. 2. Computer simulation flow chart

Computer Simulation Results

The simulated shift in the flatband voltage is shown in Fig. 3, 4, and 5 for device B at 300, 325, and 350 K, respectively. An exponential distribution of phosphorous vacancies is assumed with a surface concentration, $[V_P^o]_s$, and characteristic distance, $L \sim 5$Å (9). Parameters chosen in these simulations are: $[V_P^o]_s = 4 \times 10^{18}$cm^{-3}, $\Delta S_m/k = 15.3$, and $E_a = 1.1$ eV, respectively. Note that $[V_P^o]_s$ and E_a are experimentally deduced values (9) and that $\Delta S_m/k$ is the only adjustable parameter in the simulation.

The results of the simulation show reasonable agreement with the experimentally determined temperature dependence of the flatband shift. The linear relationship between $\sqrt{\Delta V_{FB}}$ *vs.* log t is also found in the simulation, although the slopes and intercepts of the simulated curves deviate slightly from those observed experimentally. Note that a linear relationship between $\sqrt{\Delta V_{FB}}$ *vs.* log t is observed experimentally, independent of which DCD mech-

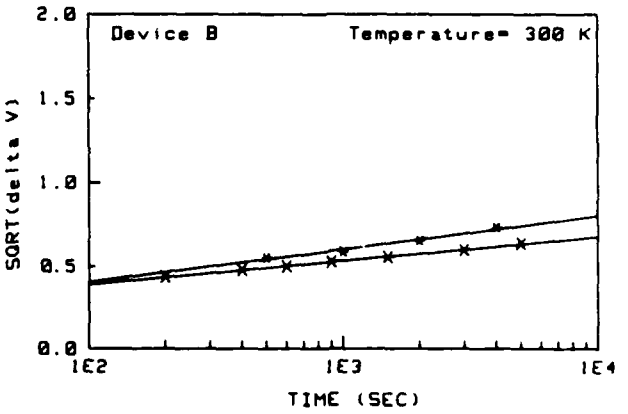


Fig. 3. Computer simulation of the flatband shift for device B at 300 K. *, Experimental data, x, simulated data. The straight lines are least squares fits to the data. $[V_P]$, 4 · 10^{18}cm^{-3}; E_a - 1 1 eV; τ, 0 108s

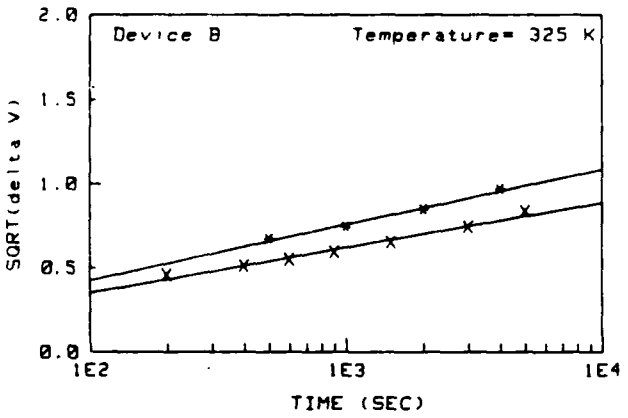


Fig. 4. Computer simulation of the flatband shift for device B at 325 K. *, Experimental data, x, simulated data. The straight lines are least squares fits to the data. $|V_P|$, 4×10^{15}cm^{-3}; E_a, 1.1 eV; τ, 0.041s.

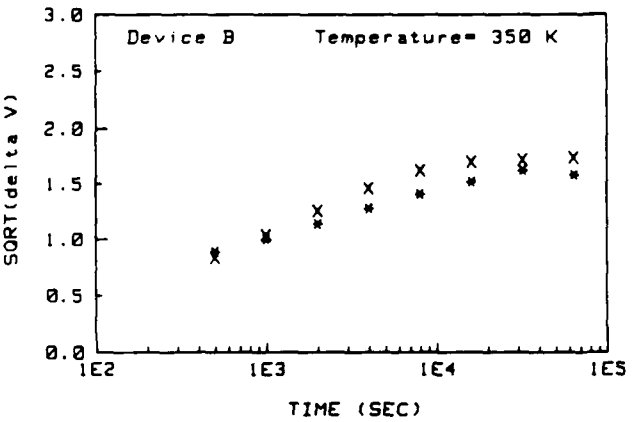


Fig. 6. Computer simulation of the flatband shift at long bias times for device B at 350K. *, Experimental data, x, simulated data.

anism dominates; this linear relationship is obtained at low temperatures and short bias times, where native oxide trapping is most important, as well as at high temperatures and long bias times, where PVNNH dominates.

The fact that both physical mechanisms are characterized by the same linear relationship between $\sqrt{\Delta V_{FB}}$ *vs.* log t is no coincidence: charge trapping and the interface electrostatics are described in a similar manner for both physical mechanisms of DCD. The trapped electron concentration for PVNNH is characterized by the first-order differential equation indicated in Eq. [17]. The analogous differential equation which accounts for electron trapping in the native oxide is (13)

$$\frac{dN_{tr}}{dt} = s(x)\, v\, \{(N_{Ta} - N_{tr})[e^-] - n_1 N_{tr}\} \qquad [27]$$

where $s(x)$ is the capture cross section of the oxide trap located at a distance x into the oxide, v is the thermal velocity, N_{ta} is the total trap density, and n_1 is the equilibrium electron concentration when the Fermi level is at the trap energy. The similarity between Eq. [17] and [27] is obvious. Note also the similarity between the second and third terms in Eq. [19] which describe the interface electrostatics and are associated with electron trapping by PVNNH and the native oxide, respectively.

The value $\Delta S_m/k$ in Eq. [2.10] was chosen to be 15.3 in the simulation, which Van Vechten and Wager (5) estimated from

$$\Delta S_m/k = \ln (8D_o/\gamma a^2 \nu_o) \qquad [28]$$

where D_o is the pre-exponential factor of the diffusion constant for In self-diffusion in InP, a = 5.869A is the lattice parameter, γ is a geometrical constant equal to 1.0, and ν_o is the attempt frequency. Using $D_o = 1 \times 10^5$cm/s as reported by Goldstein (14), gives $\Delta S_m/k$ = 17.5. But determination of D_o from experiment is very difficult and there is always uncertainty in this value. For this reason $\Delta S_m/k$ is considered an adjustable parameter (the only adjustable parameter) in the simulation.

A comparison of the experimental and simulated flatband shift for very long bias times at 350 K is shown in Fig. 6. The agreement is quite good, especially since some of the deviation between these curves can be attributed to measurement artifacts when the flatband shift is near saturation, as discussed in the previous paper (9). The simulation curve exhibits the same saturation characteristics as observed experimentally. It is clear from this simulation that saturation is indeed due to PVNNH [see the previous paper, Ref. (9), also] since the flatband shift due to native oxide trapping was estimated from an extrapolation of the low temperature experimental curves and this could not lead to saturation.

It is possible to obtain a better fit to the experimental data by assuming more complex vacancy concentration profiles or by treating $[V_P^o]_s$ and E_a as freely adjustable parameters. This parameter estimation approach was avoided since the intent of this work was to minimize the number of adjustable parameters, use experimentally determined parameters, and employ the most simple and realistic physical models. Furthermore, the agreement between the experimental and simulated curves is quite good and further variation in parameters will lead to no new physical insight.

The computer simulation also indicated that with a larger gate bias, the flatband shift is larger. A variation of the InP doping concentration does not, however, have a strong effect on the amount of flatband shift.

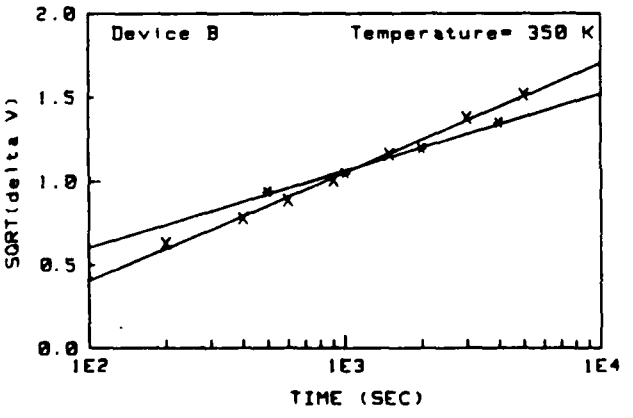


Fig. 5. Computer simulation of the flatband shift for device B at 350 K. *, Experimental data, x, simulated data. The straight lines are least squares fits to the data. $|V_P|$, $= 4 \times 10^{15}$; E_a, 1.1 eV; τ, 0.00248s.

Conclusions

A computer simulation of the flatband shift as a function of bias stress time based on an analysis of the three-step kinetics of the PVNNH defect reaction sequence agrees quite well with the experimental results. We regard this quantitative agreement as further evidence that nearest neighbor hopping is a dominant mode of atomic migration in III-V semiconductors.

Acknowledgments

We wish to thank John Arthur, Thomas Dobson, Shun Lin, S. J. Prasad, and George Pubanz for useful discussions. This work was supported in part by the Air Force Office of Scientific Research under Contract AFOSR 86-0309 and by a grant from the Murdock Foundation.

Manuscript submitted July 28, 1988; revised manuscript received Feb. 21, 1988. This was Paper 1227 presented at the Honolulu, Hawaii, Meeting of the Society, Oct. 18-23, 1987.

Oregon State University assisted in meeting the publication costs of this article.

REFERENCES

1. J. A. Van Vechten, *This Journal*, **122**, 419, 423, 1558 (1975).

2. J. A. Van Vechten, *J. Electron. Mater.*, **4**, 1159 (1975).
3. J. A. Van Vechten, *Czech. J. Phys. B*, **30**, 388 (1980).
4. J. A. Van Vechten, in "Handbook on Semiconductors," Vol. 3, S. P. Keller, Editor, Chap. 1, North-Holland Pub. Co., Amsterdam (1980).
5. J. A. Van Vechten and J. F. Wager, *J. Appl. Phys.*, **57**, 1956 (1985).
6. J. A. Van Vechten and J. F. Wager, *Phys. Rev. B*, **32**, 5259 (1985).
7. W. D. Laidig, N. Holonyak, Jr., M. D. Camras, K. Hess, J. J. Coleman, P. D. Dapkus, and J. Bardeen, *Appl. Phys. Lett.*, **38**, 766 (1981).
8. M. D. Camras, N. Holonyak, Jr., K. Hess, M. J. Ludowise, W. T. Dietz, and C. R. Lewis, *ibid.*, **42**, 185 (1983).
9. M. T. Juang, J. F. Wager, and J. A. Van Vechten, *This Journal*, **135**, 2019 (1988).
10. H. Temkin, B. V. Dutt, and W. A. Bonner, *Appl. Phys. Lett.*, **38**, 431 (1981).
11. J. P. Buisson, R. E. Allen, and J. D. Dow, *Solid State Commun.*, **43**, 833 (1982).
12. S. M. Sze, "Physics of Semiconductor Devices," Chap. 7, John Wiley & Sons, Inc., New York, (1981).
13. S. J. Prasad, Ph.D. Thesis, Oregon State University, Corvallis (1985).
14. B. Goldstein, *Phys. Rev.*, **121**, 1305 (1961).

Reprinted from Journal of the Electrochemical Society
Vol. 135, No. 8, August 1988
Printed in U.S.A.

Phosphorous Vacancy Nearest Neighbor Hopping Induced Instabilities in InP Capacitors

I. Experimental

M. T. Juang, J. F. Wager, and J. A. Van Vechten*

Department of Electrical and Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, Oregon 97331

ABSTRACT

Variable temperature bias-stress measurements were performed on n-type InP MIS capacitors. Two distinct activation energies at 40-50 meV and 1.1-1.2 eV were obtained over a temperature range of 100-350 K. These energies are consistent with the instability mechanisms of thermionic tunneling into native oxide traps and phosphorous vacancy nearest neighbor hopping (PVNNH). The estimated fraction of shift in these particular samples due to PVNNH varies both with stress time and with temperature from about 20% for short times at 300 K to about 80% for long times at 350 K.

The drain current of an InP metal insulator semiconductor field effect transistor (MISFET) is often observed to decrease as a function of time after the application of a positive gate bias which induces an accumulation of electrons in the channel. Various models have been proposed for this drain current drift (DCD) phenomena [see Ref. (1) for a critical review of proposed DCD models].

In this study, we have employed variable temperature bias-stress measurements (see the section on Experimental Procedure for a description of this technique) of InP MIS capacitors in order to determine the dominant DCD mechanisms from an analysis of the activation energy of the flatband shift. There are two advantages inherent in bias-stress measurements of MIS capacitors compared to DCD measurements of InP MISFET's. First, fabrication of the MIS capacitor requires fewer processing steps so that the interface chemistry can be precisely controlled and the electrical instabilities may be correlated to the interface chemistry. A second advantage is that the flatband shift depends only on the carrier density in the channel while drain current instabilities depend on both the carrier density and carrier mobility. Thus, interpretation of bias stress measurements is more direct than that of DCD.

Two distinct activation energies at 40-50 meV and 1.1-1.2 eV were obtained from variable temperature bias-stress measurements over a temperature range of 100-350 K. The 40-50 meV activation energy dominates the flatband shift at low temperatures and is consistent with thermally activated tunneling of electrons from the InP conduction band into a discrete trap in the native oxide. An activation energy of 1.2 eV was predicted (2) for phosphorous vacancy nearest neighbor hopping (PVNNH) in which the channel electrons are captured by shallow acceptors that are created by the hopping of an In atom into a phosphorous vacancy.

PVNNH leads to DCD in the following manner. Consider an InP MISFET which has processing-induced P vacancies in the channel region under its gate. Nearest neighbor hopping of an In atom into the P vacancy is described by the following defect reaction

$$In_{In} + V_P^{+} + 4e^{-} \rightarrow V_{In} \quad In_P^{+} \qquad [1]$$

*Electrochemical Society Active Member.

Application of the law of mass action to this defect reaction results in

$$\frac{[V_{In}\ In_P^{\ 2}]}{[V_P^{\cdot}]\,[In_{In}]} \alpha [e^{\cdot}]^4 \qquad [2]$$

The application of a positive gate bias to an InP MIS capacitor or MISFET induces an accumulation of electrons in the channel and hence an increase in $[e^{\cdot}]$ which shifts the equilibrium to the right-hand side of the defect reaction. Four of the accumulated channel electrons are captured for every P vacancy annihilated. DCD is attributed to the loss of these electrons from the channel.

Experimental Procedure

Undoped (n-type $\approx 4 \times 10^{15}/cm^3$), (100) InP wafers were used in this study. The InP surface preparation consisted of degreasing boils in tetrachloroethylene, followed by rinses in methanol/acetone and methanol. The InP surface was then prepared in one of two ways: (i) room temperature KOH/methanol—the sample was treated for 1 min in a solution of KOH in methanol (2.5g of KOH in 200 ml of methanol) which was maintained at room temperature; (ii) "hot" KOH/methanol—the sample was treated for 1 min in the KOH/methanol solution heated to 70°C. The samples were subsequently rinsed in methanol, isopropyl alcohol, and blown dry with nitrogen. A layer of approximately 850Å of plasma-enhanced chemical vapor deposited (PECVD) SiO_2 was then deposited at 300°C. An Al electrode was used as a gate contact and a soldered In contact was used as an ohmic contact. The characteristics of similar devices were reported by Krivanek *et al.* (3). The ellipsometric thickness of the native oxides reported for samples A and B are 15 and 9Å, respectively. C-V measurements indicated that sample B has higher interface state density. Room temperature bias-stress measurements were performed on approximately ten InP MIS capacitors which were subjected to different surface treatments. Variable temperature bias-stress measurements were performed only on samples A and B. The empirical activation energies of these samples were nearly identical even though the surfaces were prepared differently and exhibited different physical properties.

Following the approach of Prasad (4, 5), characterization of the trapping of electrons in InP MIS capacitors is accomplished by measuring the flatband voltage shift of the C-V curves as a function of time at different temperatures under constant positive bias. The C-V analysis was performed with an automated system based on Hewlett-Packard equipment consisting of a Model 9845 desktop computer in conjunction with a 4275A digital LCR meter, a 4140B picoammeter/dc voltage source, and a 3455A digital voltmeter. The temperature-controlled probe system used in this research was an MMR Technologies, Incorporated, low temperature microprobe system, Model LTMP-3.

The measurement is initiated by shorting the sample and heating to 375 K for 10 min, in order to reproducibly initialize the trap occupancy. The temperature is then lowered to 350 K for 10 min to achieve thermal equilibrium. An initial C-V curve, which is considered to be stress-bias free, is then obtained by sweeping at a rate of 100 mV/s from +3 to -3V. Upon completion of the sweep, a +4V bias is applied to the gate of the capacitor for the desired amount of stress time, and another C-V curve is obtained. The bias stress time is defined as the total amount of bias time applied to the sample prior to obtaining the C-V curve. The same procedure is repeated for progressively increasing stress-bias times. For example, if the desired stress-bias time is determined to be 500, 1000, 2000, and 4000s, the bias time between successive sweeps of the C-V curves is 500, 1000, and 2000s, respectively. An example of the measured C-V curves is shown in Fig. 1, which illustrates that the C-V curve shifts to the right under positive bias.

After all the desired C-V scans at 350 K are obtained, the sample is shorted and the temperature is again raised to 375 K for 1h and the value of the capacitance at zero bias returns almost completely to its initial value. The temperature is then lowered to 325 K and another positive bias C-V measurement is obtained. This procedure is continued with 25 K intervals until a temperature of 100 K is reached.

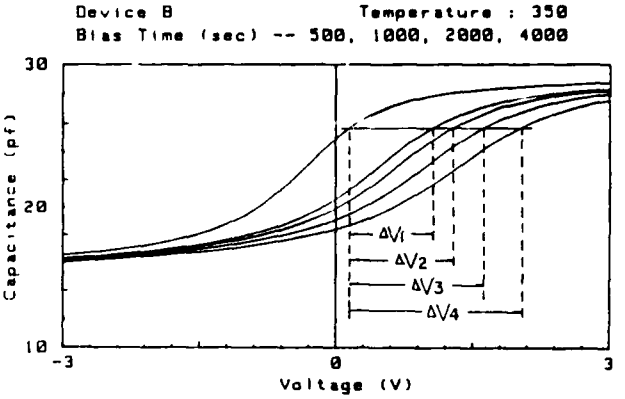


Fig. 1. Example of C-V curves measured for device B at 350 K under a 4V positive bias stress. The flatband shifts correspond to stress times of 500, 1000, 2000, and 4000s.

Experimental Results

Activation energy extraction.—As originally observed by Prasad (4, 5), a linear relationship is obtained if the square root of the flatband shift, $\sqrt{\Delta V_{FB}}$, is plotted *vs.* the logarithm of the stress time, log *t*, as shown in Fig. 2 for sample B. All of our samples at all temperatures measured have exhibited this linear behavior when plotted in this manner unless the temperature is high enough and the time is long enough that the flatband shift begins to saturate.

Activation energies associated with the mechanisms of electron trapping are determined from the slope of an Arrhenius plot; log (1/*t*) *vs.* 1000/*T*, where *T* is the temperature in K and *t* is the bias time required to achieve a predetermined amount of shift in V_{FB}, *e.g.*, 0.4V, at different temperatures. The Arrhenius curve (Fig. 3) was obtained at a flatband voltage shift 0.4V and the bias time at each temperature was interpolated or extrapolated from the experimentally observed linear relationship of $\sqrt{\Delta V_{FB}}$ *vs.* log *t*. Using a least squares fit, the slopes and intercepts can be used to extrapolate or interpolate the amount of shift at any bias time or temperature.

The low temperature activation energy of approximately 40-50 meV is consistent (7, 8) with the model of thermionic tunneling of electrons into a discrete trap in the native oxide. The high temperature (300-350 K) large activation energy (~0.45 eV) must be explained by invoking another electron trapping mechanism. If it is assumed that only two mechanisms contribute to the shift in V_{FB}, this larger activation energy is actually an effective value which is a combination of two activation energies, $\Delta E1$ and $\Delta E2$, where $\Delta E2$ is the low temperature activation energy of 40-50 meV.

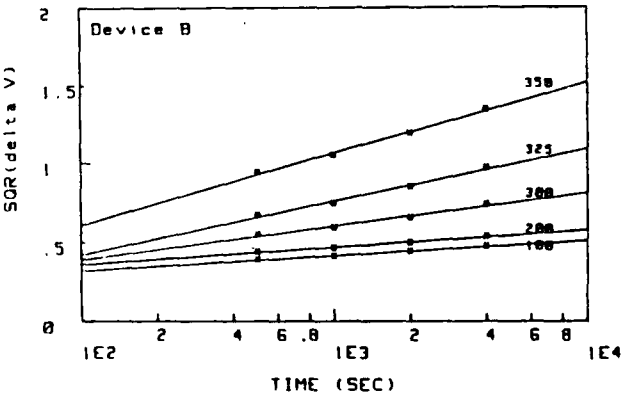


Fig. 2. Square root of the flatband shift vs. the logarithm of the stress time for device B at five different temperatures.

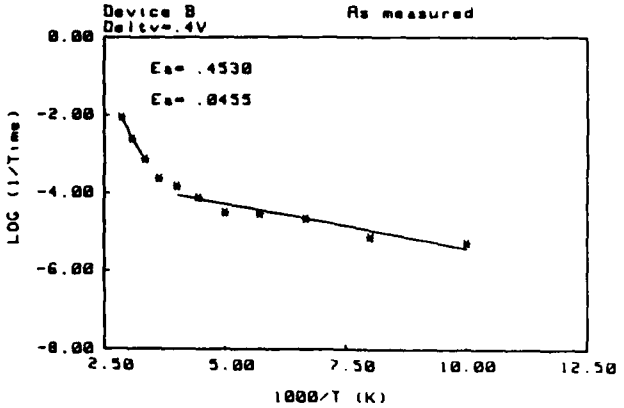


Fig. 3. Arrhenius plot of device B at $\Delta V_{fb} = 0.4V$

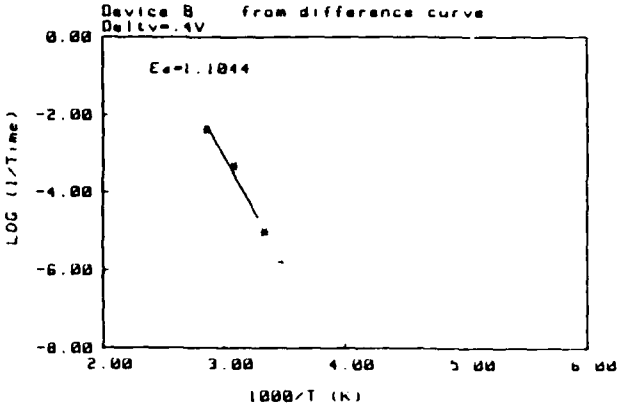


Fig. 5. Arrhenius plot of device B from the difference curves shown in Fig. 4.

In other words, in the high temperature (300-350 K) range, two electron trapping mechanisms are operational and the measured slope of the Arrhenius curve yields an effective activation energy due to the combined effects of both mechanisms. In contrast, in the low temperature (below 275 K) range, only one electron trapping mechanism with an activation energy of 40-50 meV is important. If it is assumed that these two mechanisms can be decoupled, an estimate of $\Delta E1$ can be obtained. From the Arrhenius plot, it is apparent that the mechanism with the activation energy $\Delta E1$ starts to contribute to the flatband shift at about room temperature. Thus, the amount of flatband shift due to thermionic tunneling of electrons into the native oxide can be extrapolated into the high temperature regime and the amount of flatband shift due exclusively to the second mechanism may be estimated. This is accomplished by rearranging the data between temperatures of 100-250 K into four groups according to four bias times. Four bias-stress curves are then fitted, namely, one curve for each bias time, 500, 1000, 2000, 4000s. The amount of shift for these four bias times is then extrapolated into the high temperature regime for the three temperatures, 300, 325, and 350 K.

The extrapolated amount of shift is then subtracted from the measured data, which yields a difference curve as shown in Fig. 4. This difference curve corresponds to the flatband shift when electron trapping is due exclusively to the $\Delta E1$ mechanism. $\Delta E1$ is then determined by plotting an Arrhenius curve from the difference curve. As shown in Fig. 5, the activation energy deduced in this manner is 1.1 eV for sample B which is in close agreement with that predicted (2) by the PVNNH model. A similar analysis of sample A leads to an activation energy of 1.2 eV. Since the two DCD mechanisms have been decoupled, it is now possible to estimate from the flatband shifts the relative amount of DCD that is associated with each individual mechanism. The estimated (6) fraction of the shift due to PVNNH for sample B at 300, 325, and 350 K and at various stress times is compiled in Table I. This table clearly indicates that PVNNH dominates at high temperatures and longer stress times.

Long bias time measurement.—Bias-stress measurements were performed on sample B at 300 and 350 K using very long bias times (up to 64,000s). At 300 K the $\sqrt{\Delta V_{fb}}$ vs. log t curve is linear up to stress time of 64,000s. At 350 K, however, there is a deviation from linearity at approximately 10,000s and the curve exhibits saturation near 32,000s, and a slight decrease at 64,000s as shown in Fig. 6. This decrease is probably an artifact of the measurement. After biasing for 64,000s, the 4V positive bias is removed in order to perform the C-V sweep to determine the flatband shift. If the bias is removed, even for a short time, while the flatband shift is in saturation, the amount of shift measured would be substantially reduced due to a reverse reaction that is operational during the C-V scan. Thus, at 350 K, this sample exhibits saturation in the flatband shift at a stress time of approximately 32,000s.

Although a casual inspection of Fig. 6 might lead one to conclude that there is a single saturation in which the PVNNH and thermionic tunneling mechanisms both saturate simultaneously, a careful computer-aided analysis of the data, as compiled in Table I, implies that the relative fraction of the trapping by the two mechanisms varies with both time and temperature. It is clear from the trends shown in Table I that the saturation exhibited in Fig. 6 is due to PVNNH.

Native oxide trap density and phosphorous vacancy concentration.—A rough estimate of the trap density in the native oxide and the phosphorous vacancy concentration in the InP near the interface can be made from Fig. 6. Taking the saturated flatband shift at a bias stress time of 32,000s, the native oxide trap density can be approximated as

$$N_{ox} = \frac{\Delta V_{nox} C_{ox}}{t_{nox} q} \quad [3]$$

where ΔV_{nox} is the shift in the flatband voltage due to electrons trapped in the native oxide only, which is extrapo-

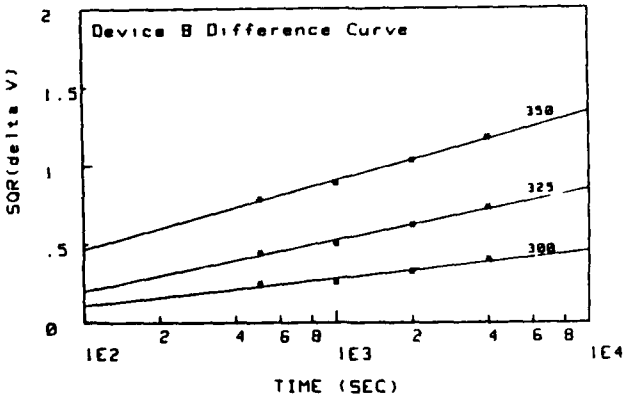


Fig. 4. Difference curve plots of the square root of the flatband shift vs. the logarithm of the stress time for device B at three different temperatures. These curves represent the flatband shift due exclusively to PVNNH.

Table I. Estimated fraction of the flatband shift due to PVNNH for sample B at various temperatures and times

T (K)	t (s)	ΔV	$\sqrt{\Delta V}$	% Total
350	500	0.6155	0.7845	70.0
	1000	0.7863	0.8867	71.7
	2000	1.0704	1.0346	75.1
	4000	1.3884	1.1783	76.5
325	500	0.1967	0.4435	43.9
	1000	0.2535	0.5035	46.2
	2000	0.3840	0.6197	53.2
	4000	0.5317	0.7292	56.8
300	500	0.0625	0.2500	20.7
	1000	0.0680	0.2608	19.5
	2000	0.1073	0.3276	25.1
	4000	0.1614	0.4017	29.7

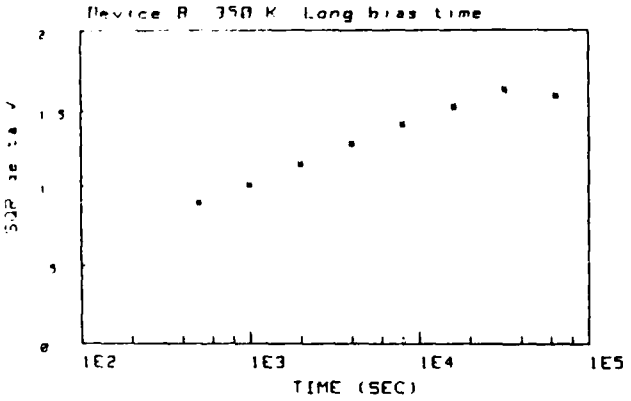


Fig. 6. Long bias-stress time measurement of device B at 350 K

lated from the low temperature data, and t_{nox} is the thickness of the native oxide. The estimated value for the native oxide trap concentration is 2.2×10^{18} cm $^{-3}$. This value compares fairly well with 3.7×10^{18} cm $^{-3}$ estimated by Lile and Taylor (8) from field-effect mobility measurements, but is larger than the value 1.5×10^{17} cm $^{-3}$ estimated by Okamura and Kobayashi (7). Some variation in the estimated oxide trap density is to be expected between various researchers because of variations in device processing and differences in the techniques used to estimate the trap density.

The phosphorous vacancy concentration at the InP surface can be estimated by assuming a simple exponential distribution of the trapped electrons in the InP accumulation layer with maximum concentration at the surface. The P-vacancy concentration is then estimated as

$$[V_p] = \frac{\Delta V_{PVNNH} C_{ox}}{3Lq} \quad [4]$$

where ΔV_{PVNNH} is the shift due only to PVNNH which is obtained by subtracting the flatband voltage shift in the native oxide from the total flatband voltage shift. The factor of three in the denominator is associated with the capture of three electrons concomitant with the In atom nearest neighbor hop (6, 10). L is the characteristic distance for the formation of phosphorous vacancies which is processing dependent. As a rough estimate of L, we assume that the creation of phosphorous vacancies in these samples arises from a collision cascade process (9) associated with the PECVD dielectric deposition and we take $L \sim 5$Å. The estimated value for the phosphorous vacancy concentration at the InP surface is 4×10^{18}/cm^3, which is a reasonable value for these samples since a phosphorous overpressure was not employed in preparing the sample and the oxide is deposited under direct plasma conditions using the plasma-enhanced chemical vapor deposition (PECVD) technique.

Discussion

Temperature dependent bias-stress measurements of InP MIS capacitors indicate the presence of two DCD mechanisms with activation energies of 40-50 meV and 1.1-1.2 eV at low and high temperatures, respectively. The low temperature activation energy is similar to that deduced (7, 8) by Okamura and Kobayashi (38 meV) and Lile and Taylor (16 meV) from DCD measurements of InP MISFET's. The low temperature DCD mechanism is due to thermally activated tunneling of electrons into discrete traps in the native oxide. The activation energy for the high temperature mechanism was predicted (2) by the PVNNH model and is thus attributed to the trapping of electrons by shallow acceptors created by the movement of a nearest-neighbor In atom into a P-vacancy.

It is well known (11-13) that deposition by indirect-PECVD, in which the InP substrate is protected from the energetic species in the active plasma, of SiO_2 in a P-overpressure results in devices with reduced DCD properties. Assuming that our devices are typical of samples prepared without a P-overpressure, the results reported here suggest that this improvement is attributable to the reduction in both the P-vacancy concentration and the density of native oxide traps. The reduction in the P-vacancy concentration is an obvious (2) consequence of equilibrium thermodynamic considerations.

It is more difficult, however, to explain the reduction in the density of native oxide traps concomitant with performing the dielectric deposition in a P-overpressure. Assuming that the P-overpressure treatment does not remove the native oxide, the P-overpressure treatment must change the nature of the native oxide (*i.e.*, the stoichiometry) in a way that reduces the native oxide trap density. Perhaps the native oxide is initially $InPO_4$ which is slightly P-depleted and the P-overpressure treatment renders the native oxide stoichiometric $InPO_4$. This interpretation is consistent with the assertion (14, 15) that native oxide traps are due to the presence of a nonstoichiometric, P-depleted native oxide. Another possibility is that a native oxide with an entirely different composition [*e.g.*, $In(PO_3)_3$ (15)] forms at the interface when the device is processed in a P-overpressure.

Conclusions

Variable temperature bias-stress measurements of InP MIS capacitors indicate the presence of two distinct DCD mechanisms: a mechanism which dominates the flatband shift at low temperatures (100-250 K) with an activation energy of 40-50 meV which is consistent with thermally activated tunneling of electrons into native oxide traps; a second, high temperature (300-350 K) mechanism with an activation energy of 1.1-1.2 eV is attributed to PVNNH. The estimated fraction of the flatband shift varies as a function of bias stress time and temperature. Approximately 20% of the flatband shift is attributed to PVNNH at 300 K for a stress time of 500s while it is almost 80% at 350 K and 4000s. The P-vacancy concentration and the density of native oxide traps were estimated at 4×10^{18} and 2.2×10^{18} cm $^{-3}$, respectively, from the saturated shift in the flatband at 350 K.

The activation energy for PVNNH was predicted (2) to be 1.2 eV, in good agreement with the empirical value reported here. This result provides additional experimental verification that nearest neighbor hopping is an important mode of atomic diffusion in III-V semiconductors (2, 16, 17).

Further work should be undertaken to perform variable temperature bias-stress measurements of InP MIS capacitors prepared in a P-overpressure to assess the extent of the reduction of the DCD and to determine the extent to which each of the two mechanisms are reduced. Surface analysis of samples prepared in a P-overpressure should provide the means of understanding how the native oxide trap density is affected by the P-overpressure treatment.

Acknowledgments

We wish to thank Joe Ryan for writing the original version of the data acquisition software and S. J. Prasad for several useful discussions. Samples used in this study were supplied by Marion Clark of Hughes Research Laboratories. This work was supported in part by the Air Force Office of Scientific Research under contract AFOSR 86-0309 and by a grant from the Murdock Foundation.

Manuscript submitted July 28, 1987; revised manuscript received Feb. 21, 1988. This was Paper 406 presented at the Honolulu, Hawaii, Meeting of the Society, Oct. 18-23, 1987.

REFERENCES

1. J. F. Wager, S. J. T. Owen, and S. J. Prasad, *This Journal*, **134**, 160 (1987).
2. J. A. Van Vechten and J. F. Wager, *J. Appl. Phys.*, **57**, 1956 (1985).
3. O. L. Krivanek, Z. Liliental, J. F. Wager, R. G. Gann, S. M. Goodnick, and C. W. Wilmsen, *J. Vac. Sci. Technol. B*, **3**, 1081 (1985).
4. S. J. Prasad, Ph.D. Thesis, Oregon State University, Corvallis (1985).
5. S. J. Prasad and S. J. T. Owen, Paper 168 presented at

the Electrochemical Society Meeting, Las Vegas, NV, Oct. 13-18, 1985.
6. M. T. Juang, M.S. Thesis, Oregon State University, Corvallis (1987).
7. M. Okamura and T. Kobayashi, *Jpn. J. Appl. Phys.*, **19**, 2143 (1980).
8. D. L. Lile and M. J. Taylor, *J. Appl. Phys.*, **54**, 260 (1983)
9. H. H. Anderson, *Appl. Phus.*, **18**, 131 (1979).
10. M. T. Juang, J. F. Wager, and J. A. Van Vechten, *This Journal*, **135**, 2023 (1988).
11. K. P. Pande and D. Gutierrez, *Appl. Phys. Lett.*, **46**, 416 (1985).
12. C. R. Zeisse, Paper 180 presented at The Electrochemical Society Meeting, Las Vegas, NV, Oct. 13-18, 1985.
13. W. Kulisch and R. Kassing, *J. Vac. Sci. Technol. B*. **5**, 523 (1987).
14. C. W. Wilmsen, J. F. Wager, K. M. Geib, T. Hwang, and M. Fathipour, *Thin Solid Films*, **103**, 47 (1983).
15. G. Hollinger, E. Bergignat, J. Joseph, and Y. Robach, *J. Vac. Sci. Technol. A*, **3**, 2082 (1985).
16. J. W. Van Vechten, in "Handbook of Semiconductors," Vol. 3, S. P. Keller, Editor, Chap. 1, North-Holland Pub. Co., Amsterdam (1980).
17. J. A. Van Vechten and J. F. Wager, *Phys. Rev. B*, **32**, 5259 (1985).

Atomic model for the EL0 defect in GaAs

J. F. Wager and J. A. Van Vechten
Department of Electrical and Computer Engineering, Center for Advanced Materials Research, Oregon State University, Corvallis, Oregon 97331

(Received 18 May 1987; accepted for publication 6 August 1986)

We propose atomic models for the stable and metastable configurations of the bi-stable, oxygen-related defect in GaAs known as EL0. They consist of an O on As site donor as a nearest neighbor to a divacancy for the stable configuration and an O on Ga site acceptor between two As vacancies for the metastable configuration. We discuss analogies and distinctions between EL0 and EL2, another important bi-stable complex. We suggest that this complex is prominent in O-containing GaAs because divacancies migrate relatively rapidly and so are prone to formation of complexes.

INTRODUCTION

Bi-stable defects, i.e., those which switch reversibly and readily between configurations in response to optical, electrical, or thermal stimulation, have been actively studied in GaAs and other semiconductors for several years.[1-10] Two of the most widely studied and important are EL0 and EL2 in GaAs. EL0 is generally regarded to contain at least one O impurity as part of the defect complex; EL2 is now widely regarded as not containing O. However, in several respects the two are very similar. The activation energies for electron emission at the complexes that have been deduced from deep-level transient-spectroscopy (DLTS) measurements for EL2 and for EL0 are virtually identical,[2] although the capture cross section for EL0 is approximately four times larger than for EL2. Both EL2 and EL0 exhibit persistent photocapacitance quenching (PPCQ), but the spectral efficiency is different[3] in the photon energy range 1.2–1.45 eV. EL2 and EL0 also have slightly different photoionization thresholds.[1]

It should be noted that O is ubiquitous in GaAs,[11] particularly that which has been grown by the liquid-encapsulated Czochralski (LEC) technique, for which, as in the case of Czochralski-grown Si, oxygen is leached from the encapsulating material. However, vendors often neglect to list O among the impurities present in their product and scientists sometimes neglect to measure it or note it in their reports. This makes it difficult to exclude O from playing a role in the structure or in the formation of EL2. It sometimes also leads to confusion of the two in observations.

There is now persuasive evidence[1-4] that EL2 is not a unique defect complex but a family of related complexes. We have proposed an atomic model[8] for the "head of the EL2 family," the core which we believe is present in all members; the other members of the family have one or more antisite pairs or extrinsic atoms added to the complex in a variety of ways. The stable and metastable configurations of EL2, which by convention are denoted as O and O* (not to be confused with atoms of oxygen), according to that model, are illustrated in Fig. 1. It must be noted that Zou *et al.* deduced that EL2 is thermodynamically equivalent to an As on Ga site-antisite defect plus a divacancy, as in our model, from a series[9] of studies of the variation of EL2 concentrations as functions of growth conditions beginning in 1981. Our model is an elaboration upon theirs which proposes specific atomic configurations, transition mechanisms, and explanations for a range of electronic, optical, and positron data in terms of empirical ionization levels of the point defects.

We present an atomic model for the "head of the EL0" family (see Fig. 2); as with EL2, other members of the family are produced by adding antistructure pairs or extrinsic atoms to the head of the family. Our proposed stable configuration for EL0 [Fig. 2(a)] differs from that for EL2 only in that the near midgap donor originates from O_{As} instead of from As_{Ga} and that it occupies a nearest-neighbor site to the V_{Ga}^{-} member of a divacancy, rather than a second nearest-neighbor site. In its metastable configuration we propose that EL0 has two As vacancies, V_{As}'s, separated by an O_{Ga}, Fig. 2(b). We propose that the transformation between these two configurations occurs simply by nearest-neighbor hopping.

Note that PPCQ measurements[10] for typical semi-insulating (SI) samples imply that the charge state of the complex is the same in both the metastable and the stable configurations. (The same is true for EL2.) As our model of the

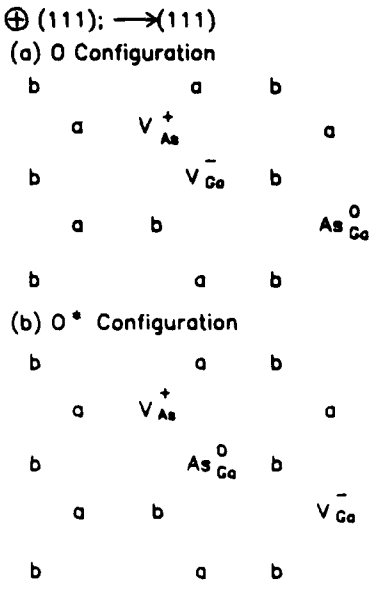


FIG. 1. (a) The stable and (b) metastable configurations of EL2 according to Ref. 8. Actually, Fig. 1 of Ref. 8 shows the As_{Ga} as a singly charged positive ion in both the stable and the metastable configurations. In fact, for most commercial semi-insulating GaAs, only a small fraction, e.g., 2%, of the As_{Ga} are positively charged at room temperature or below while the large majority are neutral.

 0021-8979/87/224192-04$02.40

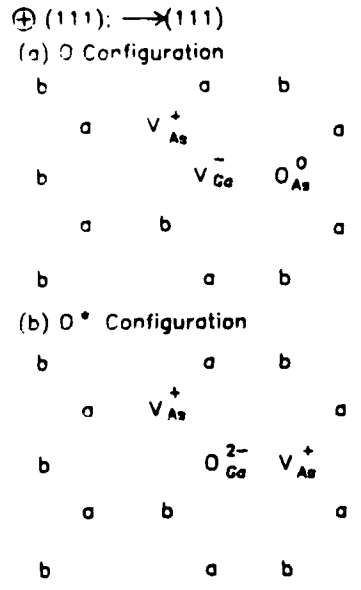


FIG. 2. (a) The stable and (b) metastable configurations of EL0 according to the model proposed herein. Ionization levels are indicated for the case of SI GaAs samples with the Fermi level near midgap.

stable configuration is neutral for these samples, we attribute an ionization state of 2^- to the O_{Ga} in the metastable configuration. Thus, we assume the O achieves its normal ionization state, and becomes a double acceptor, when it lies between two V_{As}'s. As a rigorous calculation of the ionization levels of such a complex seems to be beyond the state of the art at present,[12,13] we must rely on physical intuition, and on chemical experience, on this admittedly central point.

One should note that, unlike EL2, whose concentration varies as the square of the As activity during crystal growth,[9] our model of EL0 implies its concentration should be independent of the host-atom stoichiometry and linearly proportional to the oxygen atomic activity during crystal growth. We know of no set of experiments that could provide the data needed to test this implication.

THE MODEL

We begin to support our model by considering photoluminescence (PL) measurements.[14–20] Three broad PL bands at 0.64, 0.68, and 0.77 eV are observed at 4.2 K in undoped, SI GaAs (which, as noted above, invariably contains a substantial but generally unspecified concentration of O). All three bands exhibit persistent photoluminescence quenching (PPLQ) and have been attributed to EL2, or to EL0, or both, in the literature.

The 0.64-eV band was originally attributed to the presence of oxygen.[14,15] Although this interpretation has been questioned,[16] recent work[1,17,18] supports the original identification and an attribution to EL0. The excitation-recombination mechanism associated[19] with the 0.64-eV PL emission is illustrated in Fig. 3, which is derived from Ref. 19. The excitation begins with the ionization of a shallow acceptor (SA), which results in a bound hole localized at the SA. The 0.64-eV PL band arises from radiative recombination of a donor-acceptor pair involving a deep donor (DD) and SA. Electron capture by the ionized DD is accomplished by nonradiative (NR) capture involving an intermediate shallow donor (SD). The involvement of the SD is inferred from photoluminescence excitation (PLE) spectroscopy.[20]

This interpretation of the 0.64-eV band PL emission is completely consistent with our model of EL0. As in Ref. 8, we attribute the SA to the V_{Ga} and the SD to the V_{As}. We attribute the DD to O_{As}. The identification of O_{As} as the DD is supported by theoretical calculations[21,22] which conclude that the O_{As} should give rise to a midgap donor at a similar energy as that found for EL0. The proposed configuration of the individual defects is supported by Shanabrook *et al.*, who conclude that the DD and SA exhibit a fixed spatial separation and that they, perhaps, occupy nearest-neighbor positions.[20]

The 0.68-eV PL band is attributed to a similar donor-acceptor radiative transition in EL2, for which the DD is identified as As_{Ga}, and the hole is either a free hole in the valence band[23,24] or is trapped at a SA, which has previously been attributed[25] to V_{Ga}. The 0.77-eV band is also attributed to EL2 but thought to result from electrons in the conduction band falling to the first donor level of the As_{Ga} when this has been ionized (i.e., the level is empty).[8,23,24] These observations seem to justify our conclusion that EL0, which produces the 0.64-eV PL band but neither the 0.68 nor the 0.77 bands, differs from EL2 by involving an O atom in a position and role similar to that of the antisite As_{Ga} in EL2.

When the 0.64-eV PL emission is present at all, it is typically two to three orders of magnitude more intense than either the 0.68 or the 0.77 -eV bands.[26] It is not likely that this high relative PL intensity of the 0.64-eV emission is always due to a greater concentration of EL0 relative to EL2, because this band can dominate even when the oxygen concentration [O] is quite low. Instead, it seems that the PL efficiency of EL0 must be orders of magnitude greater than that of EL2. Our model implies this because the midgap donor is closer to the divacancy for EL0, where it is the first nearest neighbor, than for EL2, where it is second nearest neighbor. There is also much more strain in the lattice about the O impurity, which is much smaller than the host atom, than about the antisite, which is essentially the same size. These mechanisms for the high PL efficiency of EL0 are analogous to those observed in GaP when a Zn-O pair form a neutral, nearest-neighbor complex with very high luminescence efficiency.[27]

We continue to support our model by comparing and

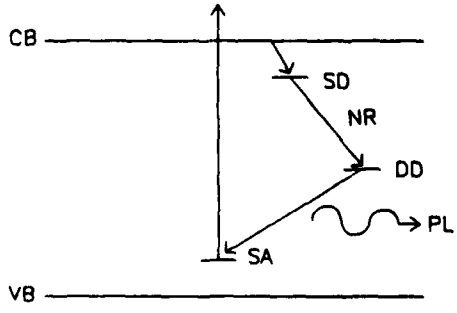


FIG. 3. Mechanism for photoluminescence of the 0.64-eV band due to EL0 proposed herein (after Ref. 19).

contrasting EL0 and EL2 with respect to the optical and electrical stimulation of the transition from the metastable, O*, state to the stable, O, state. For EL2, the electron-induced recovery, which is sometimes called[28] Auger regeneration, is known to occur[8,28] for EL2 with a thermal activation energy of 0.108 eV. Optically assisted thermal recovery from the O* to the O configuration is observed[4,29,30] above 60 K for typically about 10% of the EL2 in LEC GaAs and has a peak recovery rate at a photon energy of 0.88 eV. The fraction that exhibits this $O^* \rightarrow O$ transition varies from sample to sample; in one single case the fraction reached essentially 100%. However, the transition rate in this direction is much less than for $O \rightarrow O^*$. This distinction between most EL2's and a sample-dependent fraction of them is a part of the convincing evidence that EL2 is not a single defect but a family of related defects. We gave in Ref. 8 an explanation why, for the "head of the EL2 family," the $O^* \rightarrow O$ transition would not occur (or, better said, would have low probability) in terms of an asymmetry of electrostatic forces in the intermediate stages of the atomic reconfiguration (see Fig. 4 of Ref. 8). The same consideration would also apply to some other members of the EL2 family. However, for some of the members of the family, such as that illustrated in Fig. 4 here, the asymmetry is not so severe and the $O^* \rightarrow O$ transition would proceed at an observable rate, as is reported.

On the other hand, for EL0 the $O^* \rightarrow O$ transformation has never been reported to be induced[4] at any temperature by electron injection for any fraction of the family. Moreover, optically induced recovery from the metastable to stable configuration is not observed for the EL0 family. We attribute this distinction re $O^* \rightarrow O$ to the absence of subband-gap optical absorption at EL0's in their O* configuration, whereas such absorption does occur for EL2. Put simply, we assert that the $O^* \rightarrow O$ transition cannot be optically stimulated for EL0 because there is no optical absorption at O*'s of EL0; it can occur at (some) EL2 O*'s because they do absorb subband gap light.

To see that our model implies that this distinction should arise, we consider the O* configurations of EL2 and EL0 shown in Figs. 1(b) and 2(b). In the O* configuration of EL2, V_{As}^{+} and As_{Ga}^{0} occupy nearest-neighbor sites. The interaction between these two donors pushes the ionization level of the shallower, i.e., the V_{As}, up so that it becomes even more shallow, and that of the deeper, i.e., the As_{Ga}, down closer to the conduction band, relative to the levels of the isolated defects. However, it seems unlikely that this interaction will push either level of EL2 entirely out of the gap. In the case of EL0, the combination of two V_{As}'s, which are both relatively shallow donors, provides a potential that is even more attractive to holes and must produce a very shallow donor level; when this potential is combined with the Coulomb field of a negative O ion, the donor levels must be resonant with or very near the conduction band. Conversely, the O, which finds itself bounded by two vacancies, must be expected to attain its normal chemical state, that of an O^{2-} ion, particularly in the presence of the Coulomb potential of the V_{As}^{+}'s. Thus, the O* complex of EL0 is expected to have no net charge for any Fermi level within the GaAs band gap, or, put an equivalent other way, its ionization levels are pushed to the band edges by their mutual interaction.

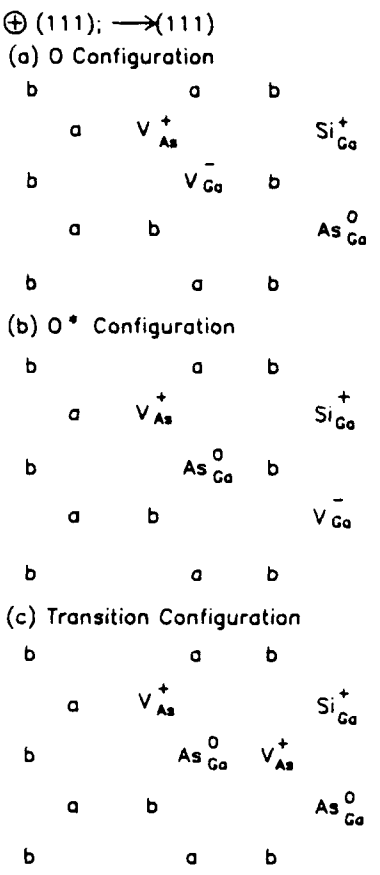


FIG. 4. Example of one of the members of the EL2 family which has a measurable probability for optically stimulated transition from the O* configuration to the O configuration (a). The stable, (b) metastable and (c) transition configurations are indicated.

In terms of the above, we expect the $O^* \rightarrow O$ optically induced transition of EL2 to proceed while for EL0 it should not because of the following: For EL2 the transition can be initiated by the photoexcitation of an electron from the neutral As_{Ga}^{0} to the conduction band,

$$As_{Ga}^{0} + h\nu \rightarrow As_{Ga}^{+} + e_c \quad . \tag{1}$$

This reaction will "turn on" a Coulombic repulsion between the V_{As} and the As_{Ga} as well as a Coulombic attraction between the V_{Ga} and the As_{Ga}. This will favor the hop of the V_{Ga} towards the As_{Ga}, which is the first step of our mechanism of the $O^* \rightarrow O$ recovery.[8] However, no analogous photoinduced defect reaction can occur in the O* configuration of EL0 because the complex has no ionization levels within the gap to serve as initial or final states for such a reaction.

A similar argument is made for the electron injection-induced $O^* \rightarrow O$ transition ("Auger recovery"). For EL2 in this case the transition is initiated by the neutralization of V_{As} by the injected electron and the activation energy is associated with a multiphonon emission process. Again this is not possible with EL0 because the V_{As} donor level is resonant with the conduction band in the O* configuration.

Finally, we note that EL0 and EL2 are so similar in their electrical and optical properties, according to our models because both families of complexes are based on the combination of a divacancy with an atom that can readily switch

from one lattice site to another and can stabilize the separated divacancy in one configuration. Si and Ge can take either lattice site but would have no way of separating a split divacancy against reformation of the bound pair in response to the divacancy binding energy, which is of order 1 eV or more.[31,32] The effect of complex formation on the ionization levels of the constituent point defects is rather moderate for reasons that are reviewed[33] by Dow. Thus, the effect of the deep level donor As_{Ga} in EL2 on the divacancy levels is not very different from the effect of the deep donor O_{As} in EL0. Moreover, it is natural that divacancies should be prominent among the defect complexes found in GaAs and other III-V semiconductors, because: (i) the large binding energy of divacancies compensates the relatively small energy of formation of antisite defects, so that the reaction

$$V_{Ga} + V_{Ga} \rightarrow V_{Ga}V_{As} + As_{Ga} \qquad (2)$$

is favorable, as asserted[9] by Zou *et al.*; and (ii) once the divacancy has formed, it is much more mobile than are single vacancies because it can move by nearest-neighbor hopping by passing atoms through the divacancy and thus avoid the 11-step process, involving the creation of three pairs of antisite defects, that is required for single-vacancy migration by nearest-neighbor hopping.[34,35]

ACKNOWLEDGMENTS

We wish to thank Professor Phil W. Yu and Professor Ralph Bray for useful discussions. This work was supported in part by the Air Force Office of Scientific Research under Contract No. AFOSR 86-0309.

[1]M. Taniguchi and T. Ikoma, Inst. Phys. Conf. Ser. **65**, 65 (1982).
[2]J. Lagowski, D. G. Lin, T. Aoyama, and H. C. Gatos, Appl. Phys. Lett. **44**, 336 (1984).
[3]M. Taniguchi and T. Ikoma, Appl. Phys. Lett. **45**, 69 (1984).
[4]Y. Mochizuki and T. Ikoma, Jpn. J. Appl. Phys. **24**, L895 (1985).
[5]J. A. Van Vechten and J. F. Wager, J. Appl. Phys. **57**, 1956 (1985).
[6]J. A. Van Vechten, Mater. Res. Soc. Symp. Proc. **46**, 83 (1985).
[7]J. F. Wager and J. A. Van Vechten, Phys. Rev. B **32**, 5251 (1985).
[8]J. F. Wager and J. A. Van Vechten, Phys. Rev. B **35**, 2330 (1987).
[9]Y. Zou, J. Zhou, Y. Lu, K. Wang, B. Hu, B. Lu, C. Li, L. Li, J. Shao, and C. Sheng, J. Electron. Mater. **14a**, 1021 (1985), and references therein.
[10]G. Vincent, D. Bois, and A. Chantre, J. Appl. Phys. **53**, 3643 (1982).
[11]R. K. Willardson and A. C. Beer, Eds., *Semiconductors and Semimetals* (Academic, Orlando, FL, 1984), Vol. 20, pp. 24, 25, 100–102.
[12]W. Hanke, Th. Golzer, and H. J. Mattausch, Solid State Commun. **51**, 23 (1984); G. Strinati, H. J. Mattaush, and W. Hanke, Phys. Rev. Lett. **45**, 290 (1980).
[13]M. Schluter (private communication).
[14]W. J. Turner, G. D. Pettit, and N. G. Ainsley, J. Appl. Phys. **34**, 3274 (1963).
[15]P. W. Yu, Solid State Commun. **32**, 1111 (1979).
[16]A. Mircea-Roussel and S. Makram-Ebeid, Appl. Phys. Lett. **38**, 1007 (1981).
[17]P. W. Yu and D. C. Walters, Appl. Phys. Lett. **41**, 863 (1982).
[18]T. Kazuno, Y. Sawaoa, and T. Yokoyama, Jpn. J. Appl. Phys. **25**, L878 (1986).
[19]D. Paget and P. B. Klein, Phys. Rev. B **34**, 971 (1986).
[20]B. V. Shanabrook, P. B. Klein, E. M. Swiggard, and S. G. Bishop, J. Appl. Phys. **54**, 336 (1983).
[21]M. Jaros, Adv. Phys. **29**, 409 (1980).
[22]A. Fazzio, L. M. Breseanint, M. J. Caldas, and J. R. Leite, J. Phys. C **12**, L831 (1979).
[23]P. W. Yu, Appl. Phys. Lett. **44**, 330 (1984).
[24]P. W. Yu, Phys. Rev. B. **31**, 8259 (1985).
[25]P. W. Yu, Solid State Commun. **43**, 953 (1982).
[26]P. W. Yu (private communication).
[27]H. Kressel and J. K. Butler, *Semiconductor Laser and Heterojunction LED's* (Academic, New York, 1977).
[28]A. Mitonneau and A. Mircea, Solid State Commun. **30**, 1571 (1979).
[29]R. Bray, K. Wan, and J. C. Parker, Phys. Rev. Lett. **57**, 2434 (1986).
[30]J. C. Parker and R. Bray, in *Defects in Semiconductors*, edited by H. J. von Bardeleben (Trans Tech, Zurich, 1986), p. 347.
[31]J. A. Van Vechten, Phys. Rev. B **11**, 3910 (1975).
[32]J. A. Van Vechten, Phys. Rev. B **33**, 2674 (1986), and references therein.
[33]J. D. Dow, Mater. Res. Soc. Symp. Proc. **46**, 169 (1985).
[34]J. A. Van Vechten, J. Appl. Phys. **53**, 7082 (1982); compare Figs. 1 and 2.
[35]J. A. Van Vechten and J. F. Wager, Phys. Rev. B **32**, 5259 (1985).

PHYSICAL REVIEW B VOLUME 38, NUMBER 15 15 NOVEMBER 1988-II

Reply to "Comment on 'Atomic model for the *EL*2 defect in GaAs'"

J. F. Wager and J. A. Van Vechten
Center for Advanced Materials Research, Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon 97331
(Received 23 May 1988)

We explain our choice of configuration of the As-antisite–divacancy complex, which components we and Zou *et al.* (in the preceding Comment) agree to comprise the "head of the *EL*2 family," in terms of electrostatics, cohesive energy, and previous interpretation (now disputed) of spin-resonance data. We propose that the configuration proposed by Zou *et al.* probably does appear within the family of defect complexes in GaAs known as *EL*2 even if it is not the "head of the *EL*2 family."

The primary issue raised by Zou, Wang, Benakki, Goltzene, and Schwab in their Comment[1] on our proposed model[2] for the *EL*2 metastable defect complex in GaAs regards its atomic geometry, not its components. Both models consist of one As_{Ga}, one V_{Ga}, and one V_{As} for the "head of the *EL*2 family" of related defect complexes. We agreed in our proposal[2] that the thermodynamic evidence, previously given[3] by Zou *et al.*, that *EL*2 varies with stoichiometry as two Ga vacancies (V_{Ga}), or equivalently, as an As on a Ga site antisite defect (As_{Ga}) plus one V_{Ga} and one As vacancy (V_{As}), is compelling. We agreed (and still agree) this is very strong evidence against models that give a different law of mass action variation of *EL*2 with stoichiometry, such as the model[4] by Van Vechten of an As_{Ga} plus two V_{Ga}'s or the model[5] of von Bardeleben *et al.* of an As_{Ga} and an As interstitial.

However, we asserted that the V_{As} component of the stable state of the *EL*2 complex is positioned at a third nearest-neighbor site from the As_{Ga}, which thus has four As nearest neighbors, whereas Zou *et al.* assert in their Comment, and in accord with a previous publication[6] by Zou, the V_{As} is a nearest neighbor to the As_{Ga}, which thus has only three As nearest neighbors, in the stable state of *EL*2. We call this the Zou model.

We chose the geometry we did because extensive electron-paramagnetic-resonance (EPR) and electron-nuclear-double-resonance (ENDOR) measurements of *EL*2 have previously always been interpreted[7–9] to require that the As_{Ga} have four As nearest neighbors. We accepted this interpretation of the EPR and ENDOR data and our model is consistent with it. This conclusion that the As_{Ga} has four As nearest neighbors is clearly in conflict with the Zou model. Thus, in their Comment, Zou *et al.* argue for an alternative interpretation of the EPR and ENDOR data which revises this crucial conclusion.

The ENDOR data have also been interpreted to show that the As_{Ga} interacts with a fifth As atom in an antibonding [111] direction.[9] In choosing the geometry we did for the *EL*2 complex, we also noted that this conclusion may be accounted for with a greater than for perfect lattice interaction of the As_{Ga} with the normal substitutional As atom in that antibonding direction which is toward the divacancy from the antisite. We argued that this results because of the free volume that comprises the divacancy. Thus, if we take As_{Ga} to be at lattice site (1,1,1), the V_{Ga} at (3,1,3) and the V_{As} at (4,2,2), we note the As_{Ga} will have a much stronger than normal interaction with the substitutional As at lattice site (4,4,4). The Zou model has only one vacancy in that region and so does not account for this conclusion from the ENDOR data nearly as well as the configuration we chose. Zou *et al.* must also argue that this ENDOR conclusion is also faulty.

We find it difficult to believe that the authors of Refs. 7–9 are completely wrong on two such fundamental and much studied points regarding *EL*2. We also favor our model geometry from a consideration of the electrostatic interaction between the donor states of the As_{Ga} and of the V_{As}; it should be energetically favorable to have the acceptor states of the V_{Ga} between these two donors, rather than outboard of them.

On the other hand there are several aspects of the Zou model that we also find attractive and which Zou *et al.* have not mentioned. First, this geometry provides a simpler mechanism for the transformation between stable and metastable configurations of the complex; the As_{Ga} need only to hop to the nearest neighboring V_{As} to produce the two V_{Ga}'s which (we assert) is reasonable to identify as the metastable configuration. Second, identification of the metastable configuration as two acceptor V_{Ga}'s (with no compensating donor states) is additionally appealing in light of the recent experimental work[10] of Mitchel, Rea, and Yu who found that *EL*2-dominated GaAs samples exhibited type conversion from *n* to *p* type upon transformation to the metastable state whereas GaAs samples dominated by *EL*0 (the oxygen-related *EL*2 level) remained *n* type (according to our closely related atomic model[11] for *EL*0, type conversion is not expected). Another advantage of this alternative geometry, and particularly of the alternative identification of the metastable configuration, is that it provides a simple explanation, in terms of compensation, for the

activation threshold observed experimentally for ion implantation into liquid-encapsulated Czochralski, semi-insulating GaAs,[12] and for the disappearance of $EL2$ at high, n-type doping concentrations.[13]

In summary, our objection to the alternative atomic geometry proposed by Zou *et al.* is that it is inconsistent with the current interpretation of EPR and ENDOR measurements. If harmony can be achieved between these experiments and their atomic model, we believe that there are several advantages of their geometry compared to that which we originally proposed. Furthermore, as we and others have argued,[2] it is clear that $EL2$ is a family of related defect complexes, not one unique defect. It may well be that both geometries occur in concentrations that vary with the thermochemical treatment of the sample. It is also most likely that different experiments will be sensitive to these two geometries to differing degrees so that when both are present, one experiment senses one and another experiment senses another.

This work is supported by the Air Force Office of Scientific Research under Contract No. AFOSR86-0309.

[1]Zou Yuanxi, G.-Y. Wang, S. Benakki, A. Goltzene, and C. Schwab, preceding comment, Phys. Rev. B **38**, 10 953 (1988).

[2]J. F. Wager and J. A. Van Vechten, Phys. Rev. B **35**, 2330 (1987).

[3]Zou Yuanxi, J.-C. Zhou, Y. Lu, K.-Y. Wang, B.-H. Hu, C.-C. Li, L.-S. Li, and Shou Jiuan, J. Electron. Mater. **14a**, 1021 (1985).

[4]J. A. Van Vechten, J. Electrochem. Soc. **122**, 423 (1975); see also J. B. Van der Sande and E. T. Peters. J. Appl. Phys. **45**, 1298 (1974); *ibid.* **46**, 3689 (1985).

[5]H. J. von Bardeleben, D. Stievenard, J. C. Bourgoin, and A. Huber, Appl. Phys. Lett. **47**, 970 (1985).

[6]Zou Yuanxi, Mater. Lett. **5**, 203 (1987).

[7]D. M. Hofmann, B. K. Meyer, F. Lohse, and J. M. Spaeth, Phys. Rev. Lett. **53**, 1187 (1984).

[8]E. R. Weber, Mater. Res. Soc. Symp. Proc. **46**, 169 (1985).

[9]J. M. Spaeth, D. Hofmann, and B. K. Meyer, Mater. Res. Soc. Symp. Proc. **46**, 185 (1985).

[10]W. C. Mitchel, L. S. Rea, and P. W. Yu, J. Appl. Phys. (1988).

[11]J. F. Wager and J. A. Van Vechten, J. Appl. Phys. **62**, 4192 (1987).

[12]B. Odekirk and R. Koyama (unpublished).

[13]J. Lagowski, H. C. Gatos, J. M. Parsey, K. Wada, M. Kaminska, and W. Walukiewicz, Appl. Phys. Lett. **40**, 342 (1982).

Reply to "Comment on 'Atomic model for the *EL2* defect in GaAs'"

J. F. Wager and J. A. Van Vechten
Department of Electrical and Computer Engineering, Oregon State University, Corvallis, Oregon 97331-3211
(Received 30 November 1987)

After correcting misimpressions regarding our atomic model for *EL2* and that of Zou *et al.*, we show that our model is in good agreement with thermochemical and positron annihilation analysis and with the compensation mechanisms which render GaAs semi-insulating, while the alternative model of von Bardeleben *et al.* is not. We give a plausible account of other electronic, optical, and electron-nuclear double resonance data regarding *EL2*.

In their comment,[1] von Bardeleben, Bourgoin, and Stievenard (VBBS) raise several issues with regard to our atomic model[2] for the *EL2* defect in GaAs. We respond here.

First, VBBS are inaccurate saying that the subject work "took up a model originally proposed by Zou." What Zou *et al.* did[3] was to show that *EL2* is thermodynamically equivalent to two Ga vacancies, or, equivalently (and in their opinion, likely), to a divacancy plus an As-on-Ga-antisite defect. The two are obviously connected through the reaction

$$2V_{Ga} \rightarrow V_{Ga}V_{As}As_{Ga} .$$

What the subject work does is to propose a model (see Fig. 1) that is consistent with this result of extensive thermodynamic measurements and study, which prescribes the atomic configurations of *EL2* in its stable and metastable states, as well as the mechanisms by which it converts from one to another, and attempts to account for all existent electronic, optical, and positron data consistently in terms of these configurations. Zou has since proposed[4] a configuration of the same point defects that is different from ours.

Second, VBBS assert that "the only experimental support for the existence of divacancy defects stems from their observation in as-grown GaAs by positron annihilation." This assertion ignores the wealth of thermochemical data[3,4] (just mentioned) analyzed by Zou *et al.* These thermochemical data demonstrate that the *EL2* concentration varies as the square of the As activity present during crystal growth. We cited ample evidence that *EL2* must be a complex of point defects, rather than a single point-defect, and there is universal agreement that the As-antisite defect is part of *EL2*. From the law of mass action we noted that the As-antisite defect concentration itself varies as the square of the As activity and any other members of the complex must appear in concentrations independent of the As activity. This restricts possibilities to stoichiometric combinations. The simplest allowed, stoichiometric combinations are the divacancy, separated Ga and As vacancies; the Ga vacancy plus the Ga interstitial; the As vacancy plus the As interstitial; the di-interstitial; and the separated Ga and As interstitials. Beyond these one could have two divacancies, a divacancy plus a di-interstitial, etc. The thermodynamic data do

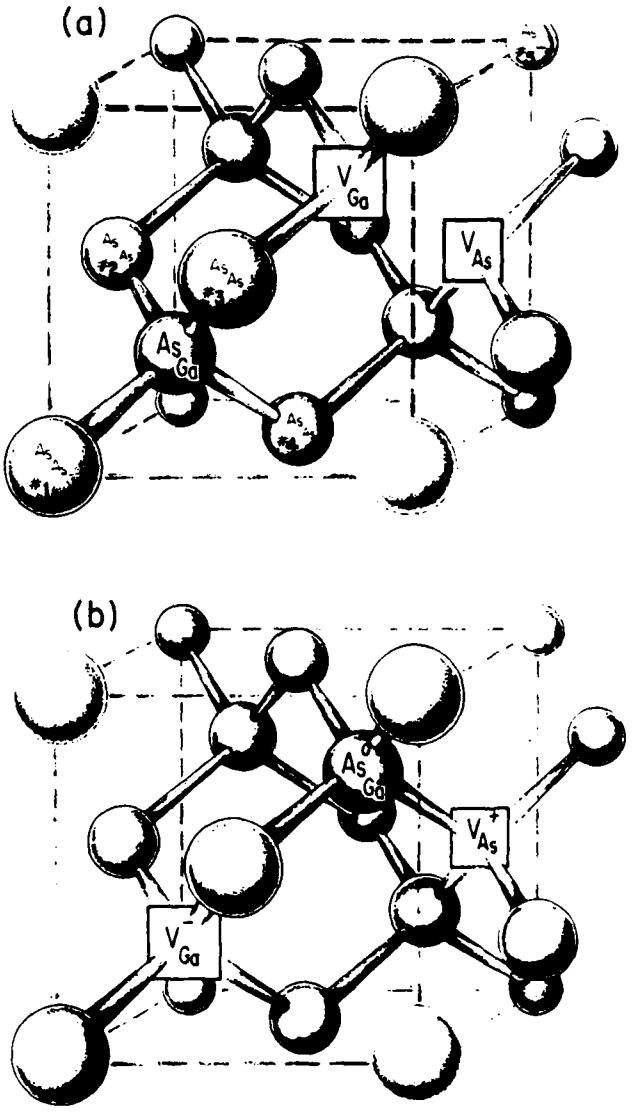


FIG. 1. The (a) stable and (b) metastable configurations of *EL2* according to Ref. 2. Actually, Fig. 1 of Ref. 2 shows As_{Ga} as a singly charged positive ion in both the stable and metastable configurations. In fact, for most commercial semi-insulating GaAs, only a small fraction, e.g., 2%, of the As_{Ga} are positively charged at room temperature or below while the vast majority are neutral.

not allow *EL2* to be, e.g., an As antisite with a single As interstitial[5] or an As antisite with a single Ga vacancy, because in either of these cases its concentration would vary as the cube of the As activity, not as the square. The positron data clearly show that there are always vacancies, i.e., single vacancies, single-vacancy complexes, and multiple-vacancy complexes, present in all GaAs samples in concentrations more than sufficient to account for all the *EL2* in the various samples and that vacancies occur on both sublattices.[6,7] Combining this demonstrated abundance of vacancies on both sublattice with any empirical determination of the enthalpy released when a vacancy and an interstitial annihilate each other,[8] it is obvious that the neither type of interstitial can be present in the subject samples in concentrations adequate to account for *EL2*. (We return to this point below.) Thus, we are left with the conclusion that *EL2* must involve stoichiometric pairs of vacancies. The only issue can be whether or not they are nearest neighbors. All experience with divacancy nearest-neighbor binding energies leads us, and Zou *et al.*, to conclude the vacancy pairs must be nearest neighbors, i.e., divacancies.

Third, VBBS argue that our proposed atomic model is inconsistent with positron annihilation measurements, since the long positron lifetime characteristic of a divacancy is only detected[6] in lightly *n*-type GaAs and anneals out at $\sim 400\,^{\circ}\mathrm{C}$ instead of above $800\,^{\circ}\mathrm{C}$ as expected for *EL2*. A simple (but incorrect) interpretation of the positron experiments is that the mean positron lifetime is determined exclusively by the abundance and size of the vacancy cavities present. An implicit assumption underlying this interpretation is that the cavities are electrically neutral. The problem with this simple interpretation is that the cavities are very often electrically charged. The cross section for positron annihilation will be substantially different for a charged defect complex compared to one that is neutral. Thus, the measured positron lifetime depends both on the size of the vacancy cavities present in the semiconductor and on the Fermi-level position which determines the charge state of these vacancy-defect complexes.

In terms of our atomic model for *EL2*, a divacancy positron lifetime will only be observed when the charge state of *EL2* is net neutral; i.e., $As^{0}_{Ga}V^{-}_{Ga}V^{+}_{As}$. This neutral complex is attained (according to the ionization energy identification as given in Ref. 2) only when GaAs is lightly *n* type, which is in accordance with the positron measurements. A divacancy positron signal will not be observed, even when a divacancy defect is present in large concentrations, when the Fermi-level position establishes a net positive charge on the defect. Thus, we believe that the positron annealing experiments are, to a large extent, determined by the Fermi-level position. Similarly, we believe that the absence of a divacancy positron lifetime in semi-insulating GaAs is associated with high-level injection and the position of the quasi-Fermi-level due to the creation of electron-hole pairs by the high-energy positrons.

Although there is presently some controversy regarding the interpretation of GaAs positron-annihilation measurements,[7] there is one unequivocal conclusion that has been established: vacancy defects are present in large concentrations ($\sim 10^{16}$–$10^{17}\ \mathrm{cm}^{-3}$) in all types of GaAs. It is exceedingly difficult for us to reconcile this fact with an As interstitial concentration equal to the *EL2* concentration that is presumed to exist. Energetic considerations imply that these As interstitials should annihilate with As vacancies to produce atoms on perfect sites and annihilate with Ga vacancies to give As_{Ga} antisites. The equilibrium energy of formation of an As-vacancy–interstitial pair has been experimentally determined[8] using differential thermal analysis by Lim, von Bardeleben, and Bourgoin to be 8.5 ± 0.5 eV, which is in good agreement with the reported[9] threshold energy of 9 eV required to create such a pair. Using the experimental value of 8.5 eV and applying the law of mass action, one can calculate the arsenic interstitial concentration, $[As_i]$, expected for liquid-encapsulated Czochralski-grown (LEC) GaAs, which is grown under near-equilibrium conditions:

$$[As_i]=\frac{N_{As}^2}{[V_{As}]}\exp[(-8.5\ \mathrm{eV})/kT]\ , \tag{1}$$

where N_{As} is the concentration of As sites, $N_{As}=2\times10^{22}/\mathrm{cm}^3$. To determine a firm upper limit of $[As_i]$, we evaluate this expression at the melting point temperature and assume a vacancy concentration at the lower limit of that inferred from positron measurements, $[V_{As}]=10^{16}/\mathrm{cm}^3$. Using these values, we conclude that

$$[As_i]=2/\mathrm{cm}^3 \ll [EL2]\ . \tag{2}$$

The unavoidable conclusion is that because of the very large energy of the As-vacancy–interstitial pair and because of the concentrations of As vacancies revealed by positron measurements, the concentration of As interstitials in unirradiated GaAs is clearly too low, by at least sixteen orders of magnitude, to account for the observed concentration of *EL2* defects.

Fourth, VBBS contend that our identification of the stable and metastable configurations is inconsistent with electron paramagnetic resonance (EPR) and electron-nuclear double resonance (ENDOR) experiments. In terms of the EPR experiment, their contention is that our proposed stable configuration of *EL2* is not paramagnetically active. We propose that, within the context of our model, *EL2* is paramagnetically active when the defect complex has a net positive charge; i.e., $As^{+}_{Ga}V^{-}_{Ga}V^{+}_{As}$. Although we identify a specific charge state to each individual vacancy, it is our view that the entire divacancy cavity is reconstructed in a way such that it is essentially neutral. If the divacancy cavity is net neutral, our model for the stable state of *EL2* is clearly paramagnetically active.

The issue of the net charge associated with the stable configuration of *EL2* is of paramount importance in the determination of the viability of any atomic model for *EL2*. Our model for *EL2* is compatible with both the two-level[10] and three-level[11,12] models which describe the compensation mechanisms which render LEC GaAs semi-insulating (SI). In their model for *EL2*,[5] it was necessary for von Bardeleben *et al.* to ascribe to As_i a

donorlike nature with an ionization energy in the upper half of the band gap to explain their experimental data. Such a model for *EL2* is clearly incompatible with the compensation requirements of LEC SI GaAs; according to the *EL2* model of von Bardeleben *et al.*, LEC GaAs would be *n* type instead of SI.

Finally, we must address the ENDOR experiments. We argued[2] that these were consistent with our model. VBBS have not countered our arguments but some have claimed such ENDOR data[13] confirms their atomic model. It is our belief that the ENDOR measurements can only be claimed to be consistent with, and not confirmation of, the model of von Bardeleben *et al.* We do not believe that this interpretation of the ENDOR data is unique or that other atomic models are not equally consistent with the ENDOR experiments. As far as we know, our proposed explanations[2] for the ENDOR measurements have not been tested for compatibility with the experimental data. We proposed that the fifth As atom which interacts with the As antisite found in the ENDOR data is the normal As atom on its undisplaced site in that unique antibonding direction which is toward the divacancy from the antisite. If, as in Fig. 1(a), the antisite is at lattice site (1,1,1) and the divacancy is on sites (3,1,3) and (4,2,2), then the fifth As atom is on site (4,4,4) in the [1,1,1] antibonding direction from the antisite. In our paper we also proposed that a major lattice distortion about the V_{Ga} portion of the divacancy cavity by one of the neighboring As atoms might also account for the ENDOR data. For example, the As atom at (2,2,4) might be displaced to nonlattice (or "interstitial") site (3,3,3), or to (2,2,2), or to any other nonlattice site on the [1,1,1] axis near the divacancy and its nominal position. The theoretical analysis of ENDOR data is apparently undergoing rapid advancement and should be instrumental in the assessment of the viability of proposed atomic models for *EL2*.

In conclusion, we have addressed the issues raised by VBBS in their comment. Our atomic model is consistent with thermochemical analysis, positron-annihilation experiments, compensation mechanisms which render GaAs semi-insulating and in our model we also attempt to account for all existent electronic and optical data. Further work is required to assess whether our model is compatible with ENDOR experiments. Our strongest objections to the model of VBBS are that it is inconsistent with the mechanism of compensation known to render GaAs semi-insulating, that it is not compatible with the thermochemical experiments of Zou *et al.*, and that it is difficult to justify the existence of such a large concentration of interstitials in the presence of the vacancy concentrations determined by positron measurements.

Note added in proof. Wang *et al.* and Goltzene and Schwab have examined[14,15] the issues of the EPR and ENDOR data and concluded that the existent data are indeed consistent with an identification of *EL2* with the complex of an As antisite with a divacancy.

ACKNOWLEDGMENT

This work is supported by the U.S. Air Force Office of Scientific Research under Contract No. AFOSR-86-0309.

[1]H. J. von Bardeleben, J. C. Bourgoin, and D. Stievenard, preceding paper, Phys. Rev. B **39**, 1966 (1989).

[2]J. F. Wager and J. A. Van Vechten, Phys. Rev. B **35**, 2330 (1987).

[3]Zou Yuanxi, Zhou Jicheng, Lu Yian, Wang Kuangyu, Hu Binghua, Lu Bingfang, Li Cuncai, Li Liansheng, and Shao Jiuan, in *Defects in Semiconductors*, edited by L. C. Kimerling and J. M. Parsey (The Metallurgical Society of the AIME, City, 1985), p. 1021, and references back to 1974 therein.

[4]Zou Yuanxi, Mater. Lett. **5**, 203 (1987).

[5]H. J. von Bardeleben, D. Stievenard, D. Deresmes, A. Huber, and J. C. Bourgoin, Phys. Rev. B **34**, 7192 (1986).

[6]S. Dannefaer and D. Kerr, J. Appl. Phys. **60**, 591 (1986).

[7]S. Dannefaer, P. Mascher, and D. Kerr, in *Proceedings of the Second International Symposium on Defect Recognition and Image Processing in III-V Compounds, Monterey, California, 1987*, edited by E. R. Weber (Elsevier, Amsterdam, 1987), p. 313.

[8]H. J. Lim, H. J. von Bardeleben, and J. C. Bourgoin, J. Appl. Phys. **62**, 2738 (1987).

[9]D. Pons and J. C. Bourgoin, Phys. Rev. Lett. **47**, 1283 (1981).

[10]C. H. Gooch, C. Hilsum, and B. R. Holeman, J. Appl. Phys. **32**, 2069 (1961).

[11]J. Blanc and L. R. Weisberg, Nature (London) **192**, 155 (1961).

[12]P. F. Linquist and W. M. Faro, in *GaAs FET Principles and Technology*, edited by J. DiLorenzo and D. Khandelwel (Artech, Dedham, MA, 1982).

[13]B. K. Meyer, D. M. Hofmann, and J. M. Spaeth, in *Defects in Semiconductors*, Vol. 10 of *Materials Science Forum*, edited by H. J. von Bardeleben (Trans Tech, Aedermannsdorf, Switzerland, 1986), p. 311.

[14]Wang Guangyu, Zou Yuanxi, S. Benakki, A. Goltzene, and C. Schwab, Phys. Rev. B **38**, 10953 (1988).

[15]A. Goltzene and C. Schwab (unpublished).

Temperature dependence of the mercury telluride-cadmium telluride band offset

K. J. Malloy[a)]
The Air Force Office of Scientific Research, Directorate of Electronic and Material Sciences, Bolling AFB, Washington, DC 20332

J. A. Van Vechten
The Center for Advanced Materials, Department of Electrical and Computer Engineering, The Oregon State University, Corvallis, Oregon 97331

(Received 10 August 1988; accepted for publication 16 December 1988)

After reviewing the experimental data on the valence-band offset for HgTe-CdTe heterojunctions, we support previous suggestions of an extreme temperature dependence for this offset with a calculation based on a bond charge model. The model predicts the T dependence of the valence-band offset to be 77% of the difference in the band-gap temperature dependence of the heterojunction constituents. In the HgTe-CdTe system, the opposite signs of the band gap T variations yield an anomalously large increase in the offset of 213 meV between 0 and 300 K.

The interest motivated by potential applications of HgTe-CdTe heterojunctions has not failed to create controversy over the fundamental properties of this semiconductor system. In particular, the alignment between the bands of the two materials has recently been the subject of much discussion. A few researchers have suggested an anomalous temperature variation of the band offset in this system,[1,2] a phenomenon often ignored in the past and in other heterojunctions. This letter briefly reviews the experimental data available on this system, particularly with regard to the temperature, and describes the results of a model explaining the origins and extent of the band offset's extreme temperature dependence.

Some of the first experimental determinations of the band offset ΔE_v, as defined in Fig. 1, came from a series of studies at 1.6 K of HgTe-CdTe superlattices. The investigators evaluated ΔE_v by fitting the magneto-optical absorption of these superlattices with the envelope function formalism.[3] Without elaborating on their fitting procedure, for a semimetallic superlattice the authors found "good agreement between experiment and theory..."[4] for $\Delta E_v = 40 \pm 10$ meV and a semiconducting superlattice gave "...acceptable agreement between experiment and the calculated transitions..."[5] for $0 < \Delta E_v < 100$ meV. Low-temperature (12 K) photoluminescence[6] and resonant Raman scattering[7] proved consistent with the absorption studies as they showed ΔE_v positive, and less than 120 meV, respectively. These experimental results also agreed with early theoretical studies.[8]

This apparent accord made the first photoemission results surprising. Conducted at room temperature, ΔE_v was found[9] to be 350 ± 60 meV. Other groups reported variations on the experiment and reached the same conclusions.[10] Although no completely satisfactory reason could be offered, Ref. 1 did point out the possibility of a temperature effect, since as mentioned above, the far infrared and resonant Raman scattering experiments were performed at low temperatures, and the photoemission experiments were conducted at room temperature.

Recently, using the unique electron tunneling transport properties at liquid-helium temperatures of a double-heterojunction structure, the authors of Ref. 11 placed an upper limit on ΔE_v of about 100 meV, fully consistent with the earlier low-temperature work. However, the same authors suggested a temperature-dependent ΔE_v when a study[2] of transport by thermionic emission in the same device fit ΔE_v with values of 245 meV at 190 K and 390 meV at 300 K.

The experimental evidence for a temperature dependent ΔE_v is by no means established since the low-temperature values of less than 100 meV arise from either the indirect outcome of superlattice studies of from the current-voltage behavior of a relatively unfamiliar device. Furthermore, recent theoretical results[12] not only predict the larger value for ΔE_v, but also fit[13] the low-temperature superlattice data with a large ΔE_v. In anticipation of further experimental studies of this offset, we wish to point out that a theoretical case *does* exist for expecting a strong temperature dependence for ΔE_v in this system.

An obvious starting point is the plausibility of some temperature variation in either the valence- or conduction-band offsets if the band gaps of the two materials forming the heterojunction have different temperature dependences. Our suspicions are immediately reinforced when we note that contrary to the conditions occurring in almost every other semiconductor pair, the band-gap temperature coefficients of HgTe and CdTe are of opposite sign.[14] To understand $d\Delta E_v/dT$, we must understand the temperature dependences of the band gaps.

It appears[15] that the thermodynamic model proposed by Heine and Van Vechten[16] is the only one to encompass both positive and negative band-gap temperature coefficients for semiconductors with the valence-band maximum at Γ. Their explanation involved using a bond charge model[17] in conjunction with a thermodynamic framework[18,19] to assess changes in phonon frequencies associated with thermal electron-hole pairs. The band-gap energy ΔE_{cv} is defined as the

[a)] Now at The Department of Electrical Engineering and Computer Science, 469 Cory Hall, The University of California, Berkeley, CA 94720.

 0003-6951/89/100937-03$01.00

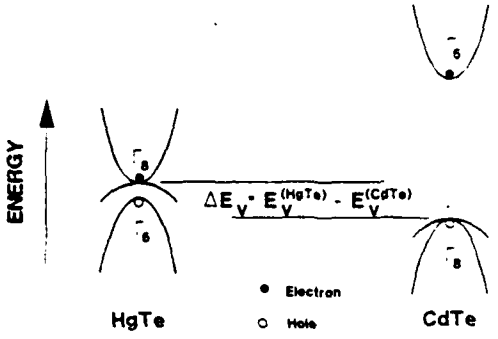


FIG. 1. Schematic E vs k diagrams for HgTe-CdTe heterojunction. The absolute energy scale is meant to point out the ability to ascribe thermodynamic free-energy changes for the relevant reactions. We use the convention assigning a negative band gap value to HgTe by calling the Γ_6 band the conduction band. The Γ_8 valence band has primarily bonding character near $k = 0$ while the Γ_6 band has antibonding character. Therefore, thermal electron-hole pairs will soften the CdTe lattice and stiffen the HgTe lattice.

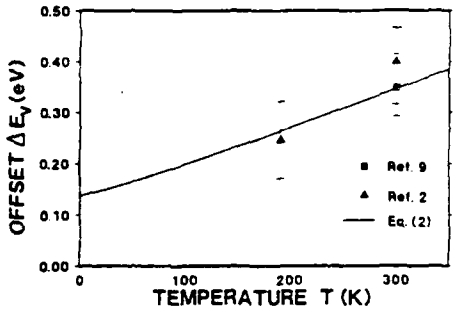


FIG. 2. Modeled empirical temperature dependence of ΔE_v using the data of Refs. 7 and 22 for the temperature variations of ΔE_{cv} and Eq. (2). Equation (2) is fixed at 300 K by the data of Ref. 9.

standard free energy of creation for the reaction $0 = e_c + h_v$, with e_c an electron in the conduction band and h_v a hole in the valence band. From Gibbs' relation,

$$\frac{d\Delta E_{cv}}{dT} = -\Delta S_{cv}, \tag{1}$$

with ΔS_{cv} the standard entropy of the reaction. This entropy arises from thermally created electron-hole pairs that lower lattice phonon frequencies when charge is removed from bonding Γ_8 states and placed into antibonding Γ_6 states, as illustrated in Fig. 1 for CdTe. The net effect is an increase in entropy due to increased occupation of lattice vibronic states and from Eq. (1), a decrease in the band gap.[17] For HgTe however, a different situation occurs. As described in Ref. 16 and shown in Fig. 1, the creation of a thermal electron-hole pair involves removing charge from the antibonding Γ_6 state and placing it in the bonding Γ_8 state, thus stiffening the lattice and increasing the phonon frequencies. The net result is an overall decrease in entropy and the required increase in ΔE_{cv} for HgTe.

Understanding the temperature dependence of the band gaps is only half the picture however, as apportioning these changes between the conduction and valence band becomes the final step to the temperature dependence of the band offsets. Recently, Van Vechten and Malloy pointed out[20] the relevance to this problem of the studies by Heine and Henry[21] of donor and acceptor isotope line shifts. Heine and Henry found in GaP that holes contribute about $3.6(\pm 1)$ times as much entropy (and are correspondingly more effective in softening the lattice phonons) as electrons. This implies[20] that in absolute energy the valence bands exhibit $3.6(\pm 1)/[3.6(\pm 1) + 1] = 77(\pm 5)\%$ of the temperature dependence of ΔE_{cv}. For ΔE_v therefore,

$$\frac{d\Delta E_v}{dT} = 0.77(\pm 0.05)\left(\frac{d\Delta E_{cv}^{(HgTe)}}{dT} - \frac{d\Delta E_{cv}^{(CdTe)}}{dT}\right). \tag{2}$$

Notice the cancellation of the offset temperature variation without the opposite temperature coefficient behavior of the CdTe and HgTe band gaps. Based on an increase in the HgTe band gap[22] between 4 and 300 K of 180 meV and a decrease in the CdTe band gap[7] of 95 meV, Eq. (2) gives an increase in the valence-band offset ΔE_v from 4 K to room temperature for the HgTe-CdTe heterojunction of $213(\pm 14)$ meV.

Figure 2 uses Eq. (2) and the empirical relationships for $\Delta E_{cv}(T)$ from Refs. 22 and 7 to summarize the temperature dependence of ΔE_v. In order to compare with experiment, we use room-temperature photoemission to fix a reference point. Our choice arises from the reproducibility of the photoemission results and the relatively simple interpretation of data. Photoemission has the advantage of being a direct optical measurement of the electronic transition encompassing ΔE_v. From Eq. (2) and the results of Ref. 9, we predict a low-temperature offset of $137(\pm 74)$ meV. Most of the low-temperature data are within our error bars.

Several comments remain to be made. A positive temperature coefficient for ΔE_{cv} also occurs for the negative gap HgSe[23] and for the positive gap alloy $Hg_{1-x}Cd_xTe$ in the range[22] $0.15 < x < 0.5$. We would expect large offset temperature variations for heterojunctions with HgTe or HgSe (or their Hg-rich alloys with other II-VIs) as one component. These considerations do suggest experiments refining aspects of Eq. (2). The ratio between hole and electron entropies of $3.6(\pm 1)$ established in Ref. 20 was for explicit experiments on GaP. While no results for other semiconductors exist, measurement of ΔE_v for a heterojunction composed of the $d\Delta E_{cv}/dT = 0$ alloy, $Hg_{0.5}Cd_{0.5}Te$, and another semiconductor should allow more insight into the constant factor in Eq. (2).

Our contribution has been establishing the importance of temperature and the origins of temperature dependence for the band offsets in the HgTe-CdTe heterojunction system. The experimental data available on this system are completely consistent with this property when the experimental temperature is taken into account. In light of several theories predicting ΔE_v for this system,[8,12] we have pointed out the utility of treating the electrons as a perturbation on the phonon system when seeking to understand temperature effects.[19] The difficulties in applying 0 K theories to all temperatures are obvious in this situation.

The work at the Oregon State University was supported

in part by grant No. AFOSR-86-0309 from the Air Force Office of Scientific Research. The authors acknowledge useful discussions with Professor T. C. McGill.

[1]J. P. Faurie, C. Hsu, and T. M. Duc, J. Vac. Sci. Technol. A **5**, 3074 (1987).
[2]D. H. Chow, J. O. McCaldin, A. R. Bonnefoi, T. C. McGill, I. K. Sou, and J. P. Faurie, Appl. Phys. Lett. **51**, 2230 (1987).
[3]G. Bastard, Phys. Rev. B **25**, 7584 (1982).
[4]Y. Guldner, G. Bastard, J. P. Vieren, M. Voos, J. P. Faurie, and A. Million, Phys. Rev. Lett. **51**, 907 (1983).
[5]J. M. Berroir, Y. Guldner, J. P. Vieren, M. Voos, and J. P. Faurie, Phys. Rev. B **34**, 891 (1986).
[6]D. J. Olego and J. P. Faurie, Phys. Rev. B **33**, 7357 (1986).
[7]D. J. Olego, J. P. Faurie, and P. M. Raccah, Phys. Rev. Lett. **55**, 328 (1985).
[8]J. O. McCaldin, T. C. McGill, and C. A. Mead, Phys. Rev. Lett. **36**, 56 (1976); W. Harrison, J. Vac. Sci. Technol. **14**, 1016 (1977).
[9]S. P. Kowalczyk, J. T. Cheung, E. A. Kraut, and R. W. Grant, Phys. Rev. Lett. **56**, 1605 (1986).
[10]J. P. Faurie, C. Hsu, and T. M. Duc, J. Vac. Sci. Technol. A **5**, 3074 (1987); C. K. Shih and W. E. Spicer, Phys. Rev. Lett. **58**, 2594 (1987).
[11]D. H. Chow, T. C. McGill, I. K. Sou, J. P. Faurie, and C. W. Nieh, Appl. Phys. Lett. **52**, 54 (1988).
[12]W. A. Harrison and J. Tersoff, J. Vac. Sci. Technol. B **4**, 1068 (1986); S.-H. Wei and A. Zunger, Phys. Rev. Lett. **59**, 144 (1987).
[13]M. Jaros, A. Zoryk, and D. Ninno, Phys. Rev. B **35**, 8277 (1987).
[14]D. Long and J. L. Schmit, in *Semiconductors and Semimetals, Vol. 5: Infrared Detectors*, edited by R. K. Willardson and A. C. Beer (Academic, New York, 1970).
[15]M. L. Cohen and D. J. Chadi, in *Handbook on Semiconductors, Vol. 2*, edited by M. Balkanski (North-Holland, Amsterdam, 1980).
[16]V. Heine and J. A. Van Vechten, Phys. Rev. B **13**, 1622 (1976).
[17]R. M. Martin, Phys. Rev. **186**, 871 (1969); W. Weber, Phys. Rev. Lett. **33**, 371 (1974).
[18]C. D. Thurmond, J. Electrochem. Soc. **122**, 1133 (1975).
[19]J. A. Van Vechten, in *Handbook on Semiconductors, Vol. 3: Materials, Properties and Preparations*, edited by S. P. Keller (North-Holland, Amsterdam, 1980).
[20]J. A. Van Vechten and K. J. Malloy, Phys Rev. B (to be published).
[21]V. Heine and C. H. Henry, Phys. Rev. B **11**, 3795 (1975).
[22]J. Chu, S. Xu, and D. Tang, Appl. Phys. Lett. **43**, 1064 (1983).
[23]M. Cardona and G. Harbeke, J. Appl. Phys. **34**, 813 (1963).

Vacancy First and Second Neighbor Hopping at a Compound Semiconductor Interface - Insights from Computer Simulation

J.A. Van Vechten and U. Schmid*
Center for Advanced Materials Research
Department of Electrical and Computer Engineering
Oregon State University, Corvallis OR 97331-3202

INTRODUCTION

As one might expect from the nature of the subject,[1,2] the array of point defects to be expected at a semiconductor heterojunction, and their rates of development, diffusion and annealing is a very complex issue. For example, the rate of interdiffusion of an AlAs/GaAs superlattice is known[3,4] to vary dramatically with sample stoichiometry, doping type, type of donor or acceptor, and level of doping. Under some circumstances the rate of diffusion is essentially the same[5-7] on the anion and cation sublattices, while in other circumstances it is much more on one[8] than on the other sublattice. Evidently, this interdiffusion process may be dominated by more than one type of point defect and the dominant mode of atomic migration may also vary.

It is plain that the point defect arrays that mediate atomic diffusion across semiconductor interfaces should be expected also to affect electron transport across, along and near these junctions.

In the face of such complexity and such relevance, one often resorts to the use of semi-empirical model theories in hopes of dramatically simplifying the analysis of raw data. One of the most dramatic of these is the Ballistic Model[9,10] (BM), which identifies the activation enthalpy of atomic migration between equivalent sites as the kinetic energy the hopping atom must have if

it completes the hop in one zone boundary phonon period. Thus, potential energies are assumed to be negligible for the short venue of the migration event. One need only know the phonon period, which in practice we estimate from the Debye temperature, θ_D, and the distance between sites. The model has been very successful[9] with elemental crystals of many types, where it asserts that the activation enthalpy for host atom A to hop to a vacant site, i.e., vacancy, on a first nearest neighbor site is

$$\Delta H_m(V_A) = \Delta H_m^{1nn}(A\colon A) = M(A)\ v^2/2 = \Delta H_k^{1nn}(A\colon A)\ , \qquad (1)$$

where M(A) is the mass of A, v is the velocity required to cover the nearest neighbor distance of the unperturbed lattice in the zone boundary phonon period, which we in practice estimate to be 0.9 k θ_D/h, and where we use ΔH_k explicitly to denote the kinetic energy component of the activation barrier when we want to consider a possible contribution from potential energies. In particular, we should consider contributions to ΔH_m from differences in the bonding configuration of impurity atoms or of antisite defects.

The BM has previously been extended[10] to pure compound semiconductors, where, for vacancy nearest neighbor hopping and similar problems, one must take account of the difference between the masses of the atoms that hop and between the sites.

An important test for the BM was the issue of the enthalpies of nearest neighbor hopping for V_{In} and for V_P in InP. The BM predicted these to be in the 1 to 4 ratio of the masses of the atoms (P and In respectively) which do the hopping while other simple models imply either that these enthalpies should be about the same or in the reverse ratio. After careful study of InP MIS capacitors with variable temperature bias-stress measurements, the activation enthalpy for V_P migration was found[11] to be 1.15 eV, in good agree-

ment with the BM predicted value. Additionally, the 0.3 eV value predicted from the BM for V_{In} hopping was identified[12] as the activation energy of transformation of the metastable M center in InP. We previously noted[10] that the BM gave a value for the activation enthalpy for second nearest neighbor hopping of O vacancies in MgO, $\Delta H_m^{2nn}(O:MgO)$, that is in good agreement with experiment and with the elaborate calculation[13] of Sangster and Stoneham, who used the HADES computer simulation code for diffusion in an ionic crystal (which is not applicable to semiconductors).

Further support for the BM has recently been found[14] in observations of the pressure dependence of the rate of vacancy diffusion in materials ranging[15-17] from Au to Si. Whereas theories that treat the activation enthalpy for vacancy migration, $\Delta H_m(V)$, as a potential energy conclude[15,16] that, because interatomic potentials increase with pressure, ΔH_m should increase rapidly. Experimentally one finds that the increase is either small or that there is actually a decrease.[15-17] Because the distance of the hop decreases with pressure often faster than the zone boundary phonon frequency increases, the kinetic energy of the hop assumed in the BM often decreases and generally does not increase nearly as much as the estimates of a potential energy barrier.

Here we explore some of the implications of the BM for problem of atomic diffusion across a semiconductor junction and the array of point defects, and point defect complexes, involved therein. We have chosen to focus on the AlAs/GaAs system because there is recent interest in and data on these junctions. Despite the great simplifications induced by the assumption of the BM, we found the problem still so complicated that we devised a microcomputer simulation program to work out the consequences of various assumptions regarding the atomic scale thermochemical parameters through many millions of

point defect migration and reaction events.[18] This program is called VIDSIM for "vacancy and interstitial diffusion simulator" and is available to anyone requesting it for the cost of the diskette and postage, $1.50.

The plan of this work is as follows. In the next section, we review the extension of the BM to atomic first and second neighbor hopping in a compound lattice. We then describe the VIDSIM program briefly. Next we present specific results for the a simulation of a 4% AlAs in GaAs alloy. We then compare with experimental data on the AlAs/GaAs system. We conclude with a brief summary.

THE BALLISTIC MODEL FOR COMPOUNDS WITH FIRST AND SECOND NEIGHBOR HOPPING

When considering the hopping of atoms to first and second neighbor sites in a compound, one has to deal with the problem of the non-equivalence of sites. A host atom will be an antisite on one and a normal host atom on the other. An impurity will be bonded differently on the two sublattices and will have different ionization levels and states.In the spirit of the BM, we have adopted the assumption that the ΔH_m's are all entirely determined by the initial configuration and not affected by the final configuration. In effect, we assume the hopping atom does not know what it will encounter on the other side of the activation barrier. This is clearly another major simplification of the problem. It is also at least consistent with the requirements of thermodynamics because the reverse reactions are dependent only on the configuration from which they start (the final state of the forward reaction), so the net reactions have the proper dependence on the difference between the two bonding configurations.

The BM implies that in a tetrahedral lattice, where the distances to the second and to the first nearest neighbor sites are in the ratio $(8/3)^{1/2}$, the

activation enthalpies for a particular atom, A, to hop from whatever its present bonding configuration in host XY may be to a vacant site at those distances to be in the ratio

$$\Delta H_k^{2nn}(A: XY)/\Delta H_k^{1nn}(A: XY) = 8/3 = 2.67 \ . \qquad (2)$$

Please note that, because this ratio is less than the 4 to 1 ratio of the masses of In and P atoms, the BM implies that in InP a P atom would more easily hop to P vacancy, V_P, on a second neighbor site than could an In atom hop to V_P on a first neighbor site. Of course, the P atom could even more easily hop to an In vacancy, V_{In}, on a first neighbor site, which would produce an antisite defect, P_{In}. It is now well known that antisite defects are common in III-V semiconductors.

We now adopt a general formula, which allows for differences in initial bonding configuration, for the activation enthalpy for atom A in host XY to hop from its present site to a vacant site on an i^{th} neighboring site

$$\Delta H_m^{inn}(A: XY) = \Delta H_k^{inn}(A: XY) + F_{NA} \sum_{j=1}^{4} E_j(A) + B_1 + \Delta H_k(vac) \qquad (3)$$

where $\Delta H_k^{inn}(A: XY)$ is the term we regard as the BM kinetic energy contribution, but which the reader may regard as any additive constant he chooses. In Eq. (3) F_{NA} is a factor that may be adjusted to allow for a deviation from the bond additivity approximation but which we have set to unity for this work; $F_{NA} = 1.0$ corresponds to that bond additivity approximation, which is that the total bonding enthalpy be the sum of enthalpies of bonds between all nearest neighbor bonded atoms with values that depend only on the two atoms participating in each bond. The $E(A)_j$ are these bond energies for the four bonds around A at its present site. The B_1 is an adjustment for vacancy neighbors

that is used to account for divacancy binding,[19-22] trivacancy binding, etc.; the subscript l denotes the number of sites, 0 to 4, about the present site of atom A which are vacant and these values are adjusted to make the multivacancy binding energies in XY come out as the user calculates or estimates from the experiment they should. (B_l depends on the chemistry of host XY but not that of hopping atom A.) $\Delta H_k(vac)$ may be regarded as an additive constant that the user may adjust to scale all activation barriers, and thus all rates of reaction, up or down uniformly. This allows him to adjust the simulated overall rate of the process but it has no effect on the configurations that come out after any number of events. For convenience in the programming, it is expressed as if it were a kinetic energy associated with the migration of the vacancy.

For our simulation of the AlGaAs alloy system doped with Si and with Te, we have used the values of $\Delta H_k^{1nn}(A, GaAs)$ shown in Table 1, which is known as the ".occ" file in VIDSIM, and scaled the values for the 2nn kinetic energy in each case by the factor 2.67 indicated at Eq. (2). We assumed the values for $E_j(A)$ shown in Table 2, which is known as the ".ocx" file in VIDSIM. These values were estimated by the methods reviewed in Ref. 2. No attempt has, as yet, been made to optimize the agreement with any part of the present data.

Note that we have here ascribed one and the same value to the E_j between any atom and a vacancy, −1 eV, and have set the B_l's equal to zero. This implies a nearest neighbor binding energy for a divacancy of +1 eV that is independent of the chemistry of the atoms surrounding the divacancy. The value is somewhat less than is observed for divacancies in Si.[19-22] It also implies the assumption that the third vacancy will bind to one of the vacant sites as strongly as did the second vacancy. While this seems to us to be a

reasonable assumption, we do not know of any empirical evidence one way or another.

We arbitrarily set $\Delta H_k(\text{vac}) = 1$ eV. This serves to cancel the −1 eV that enters Eq. (3) for single vacancy, 1nn hopping because the atom that hops then has one vacancy neighbor and so has one $E_j(\text{A–vac}) = -1$ eV term. These leaves us with a barrier that differs from the previous BM value by the sum of the other 3 E_j terms. As those terms can be either positive or negative depending on other normal or antisite neighbors and as these terms are generally small compared with the kinetic energies, the net difference is not great. As already noted, any difference plays no role in determining the configurations that will result; it will only affect the simulated rate. We have not yet determined how much the various rates are affected.

The simulation does not yet contain coulomb interaction, spatial variation of the Fermi level, or the effects of carrier generation or recombination at the defects.

For the present study, we assume no interstitials of any sort are present (except for the atom in the process of a second neighbor vacancy hop, if one wishes to regard that as an interstitial).

VIDSIM

VIDSIM[18] stands for "Vacancy and Interstitial Diffusion Simulator". The user inputs whatever assumptions he chooses for the interaction parameters among host and impurity atoms at lattice sites, antisites, bond-centered interstitial sites, hexagonal interstitial sites or tetrahedral interstitial sites and vacant sites. The user also inputs an initial configuration of the point defects in the crystal, the temperature and the "seed" for a random number generator. The program then uses a Monte Carlo method to work out the

consequent migration and reaction of the point defects through many discrete events. At present, the program is limited to diamond or zinc-blende structure lattices of no more than 2^{42} ($4.4x10^{12}$) sites with no more than 200 mobile defects (vacancies and interstitials) and about 5000 other defects (impurities and antisites) that move only by interaction with the mobile defects. When run on an IBM PC/AT with a 6 Mhz math coprocessor and 2nn hopping switched off, it typically simulates 50 discrete events (e.g., vacancy hop to first neighbor site, interstitial hop, vacancy-interstitial creation or annihilation) per second.When, as here, we allow 2nn hopping, the rate slows to about 14 events per second. The rate on other computers scales very closely with the frequency of the math coprocessor.

It should cost the user almost nothing to use VIDSIM. By agreement with the U.S. Air Force Office of Scientific Research, which funded its development, the program will be provided to anyone anywhere who requests it for $1.50 to cover the cost of one floppy disk plus postage and handling. (We are prepared to waive the charge in cases where currency transfer is difficult.) The program is designed to run in "background" with commonly available PC multitasking utilities such as SoftLogic's DoubleDos or Quaterdeck's DesqView while the computer is used for other purposes or while it would otherwise stand idle over night, etc. It requires only 225 Kbytes of the computer's RAM memory for the simulation described here. A small simulation requires as little as 120 Kbytes.

For the present study we assume a 4% alloy of AlAs in GaAs. We restrict the space of the simulation to a box 20 lattice constants on a side with period boundary conditions; atoms and vacancies hopping through the boundary in the +x direction re-enter through the -x boundary with the corresponding y and z coordinates. The box contains 64,000 lattices sites, 32,000 of each

type. For the 4% alloy we have 1280 labeled Al atoms placed initially on randomly selected sites. The doping is accomplished with 15 Si or Te atoms regularly spaced to keep them as far apart as possible. To this we have added either 4 As vacancies, V_{As}, or 4 metal vacancies, V_{Ga}.

We "annealed" this initial state for approximately $5x10^6$ events with T = 1300 K and then started to observe the varying configurations by recording the result at fixed intervals of the simulated time, usually 10^4 times the arbitrary constant determined by our choice of parameters in Tables 1 and 2. The number of events and the running time of the computer varied widely for this interval, from as few as $1x10^4$ events to more than $3x10^6$.

We also calculate one event beyond the fixed interval stopping point and recorded that configuration also. The reason for this is as follows. A thermodynamic system dwells in stable configurations longer than in those that have states of lower free energy connected to them by paths with low activation barriers, i.e., "transition states". In principle, the proper thermodynamic averages are obtained by taking "snap shots" at fixed intervals of the simulated time. However, the computer checks the total elapsed simulated time against the limit after each event. Therefore, it is much more likely to exceed the limit and record the result when passing out of the stable state to the transition state, than when passing from the transition state to the next stable state. The next hop after the fixed interval snap-shot is often the state from which the snap-shot state was derived because an atom often hops directly back again after hopping out of a stable configuration. Otherwise the next hop is often to another stable configuration. Thus, we usually find the transition state at the snap-shot and the stable configuration at the next hop.

After the initial "annealing" phase we simulated mostly with T = 1100 K. Lower T's were tried but seemed to take too long for the configurations to change significantly. We here report only T = 1100 K results.

RESULTS OF SIMULATION IN AlAs-GaAs 4% ALLOY

By comparison with a study of V_{As}'s and V_{Ga}'s in pure GaAs, we observe that the presence of the light Al atoms has a dramatic effect on the configurations assumed by our 4 simulated V_{As}. Within the initial "annealing" period they, formed a trivacancy complex which diffuses very much more rapidly than the single vacancy. Although it constantly brakes up and reforms as it diffuses, it stays together for very long simulated times. (It broke up after 50 days of running time on a 6 Mhz PC/AT. We do not yet have a large statistical base of trivacancy complexes.) In our simulation so far (about 10^8 events at T = 1100 K), the single vacancy has never joined the trivacancy. Statistics on the snap-shot and "next" state configuration of the 4 vacancies are given in Table 3. We note a distinct majority of the vacancies now occupy metal sites. This occurs for the single vacancy as well as for the trivacancy. Statistics for the antisite defects at an arbitrarily chosen snap-shot configuration are given in Table 4. We note that the Al_{As} antisite are present well beyond their 4% weight in the alloy, particularly among the antisite complexes.

For 22 observations of the trivacancy complex, one Al antisite was found as a nearest neighbor of at least one of the three vacancies 8 times. Once we found two Al on nearest neighbor site of one As vacancy which had metal vacancies as its other two nearest neighbors. These frequencies of occurrence are far beyond what would be expected were the Al atoms randomly distributed around the trivacancy. For the most common form of the trivacancy complex,

$V_aV_bV_a$, with a random distribution of Al the probability of finding any Al nearest neighbor would be 15.07% so the expectation value for 22 tries would be 1.7.

On the other hand, as expected for a random distribution, we never found an Al atom on a nearest neighbor site of the one isolated vacancy in the same 22 runs, nor did we find an Al nearest neighbor to either the one divacancy or the two isolated vacancies that migrated trough the sample for 9 runs (Run #62 to Run #70) after the initial trivacancy broke up and before the next formed.

We also note that the same labeled Al atom migrates for substantial distances with the trivacancy. Thus, the easy ability of the Al to hop to second neighbor sites, as well as to first nearest neighbor sites evidently leads to: a) the formation of trivacancies; b) the formation of isolated Al_{As}; and c) the entrainment of Al with the trivacancy. We suggest that the last effect constitutes a potentially important mechanism for the diffusion of Al over distances of many bond lengths.

When we run the corresponding experiment with 4 V_{Ga} and only the heavy Te dopant, we could obtain no multiple vacancy complexes although we ran the simulation at both T = 1300 K and at T = 1100 K for more than 10 times longer (simulated time - not computer running time) than required to form the trivacancy from the V_{As}. Statistics on the configurations are given in Tables 5 and 6. Note that there is less tendency to convert vacancy type, less super-abundance of Al next to the vacancies, and there are fewer antisites of all kinds. The overall rate of atomic diffusion is also much slower.

We note that if we had taken the alloy to be 4% P in GaAs, we would certainly have gotten the converse result, i.e., the V_{Ga}'s forming trivacancies and the V_{As} would not do anything interesting. This is because our

assumptions are quite symmetric and there is as yet no effect of the Fermi level in the simulation.

When we add Si to pure GaAs, we get divacancies whether we put in 4 V_{Ga} or 4 V_{As}. Of course, this is what one might expect from the amphoteric nature of Group IV impurities. When we added Si to the 4 % AlGaAs alloy and started 4 metal vacancies, we first got 1 divacancy, but after three runs in became a trivacancy and persists at this writing. We have now completed 5 more runs. However, we have not yet found any Al neighboring the trivacancy. Moreover, we are startled to find all 15 Si atoms remain on metal sites.

RECONCILIATION OF REAL DATA

We now attempt the reconciliation of several experimental reports regarding diffusion of Al and Ga in the AlAs/GaAs system which, on first regard might seem confusing or even contradictory. We consider the following: Tan and Goesele report[23] that for intrinsic GaAs,

$$D(\text{Ga: i-GaAs}) = 2.9\text{x}10^{8} \exp(-6\ \text{eV}/kT)\ \text{cm}^2/\text{s} \ . \tag{4}$$

Mei et al. report[3] that for p-type Zn doped AlGaAs

$$D(\text{Al: p-AlGaAs: Zn}) \ \alpha \ \exp(-1\ \text{eV}/kT) \ , \tag{5}$$

and for As rich, n-type, Te doped GaAs

$$D(\text{Al: n-AlGaAs: Te}) \ \alpha \ n^{+1} \exp(-2.9\ \text{eV}/kT) \ , \tag{6}$$

where n denotes the free electron concentration at the diffusion temperature. Mei et al. and Tan and Goesele agree that for As rich, n-type, Si doped AlGaAs

$$D(\text{Al: n-AlGaAs: Si}) \ \alpha \ n^{+3} \exp(-4\ \text{eV}/kT) \ . \tag{7}$$

On the other hand, for the case of rapid thermal annealing, RTA, of a capped sample of Si doped AlGaAs, Greiner and Gibbons report[4]

$$D(Al:\ n\text{-}AlGaAs:\ Si)\ \alpha\ \exp(-2.5\ eV/kT)\ . \qquad (8)$$

Furthermore, for the case of Ga rich Zn doping, substantially equal degrees of atomic intermixing are found on both sublattices,[5-7] while for the case of As rich Zn or of Si doping the degree of mixing is about 300 times greater on the metal sublattice than on the anion sublattice.[4,8]

At this writing we cannot claim to have clarified all the issues relating to these results with our VIDSIM simulation. We have not yet done the extensive statistical analysis of results at a series of simulation temperatures that would be required to make such a claim.What we offer here is a set of proposed explanations that, to some degree, have been inspired by insights from the VIDSIM computer experiments.

We also note that a proper explanation of these diffusion phenomena requires more than a simple proposal for a mechanism by which the atom in question might get from one site to another. For example, in the case of Si doping, one might note that Si donor – acceptor pairs are expected to form and that these allow an easy path for say a Ga vacancy to move from one side of the pair to the other:

$$V^{-}_{Ga}Si^{-}_{As}Si^{+}_{Ga} \;<==>\; Si^{+}_{Ga}V_{As}Si^{+}_{Ga} \;<==>\; Si^{+}_{Ga}Si^{-}_{As}V^{-}_{Ga}\ . \qquad (9)$$

It is plain that reaction (9) does not suffice to explain any long ranged migration of anything. It merely moves the vacancy one second nearest neighbor spacing and inverts a dipole. For any migration to occur, the vacancy must leave the Si pair, travel around the six-fold ring of the zinc-blende lattice and approach again from a different direction, or it must

be replaced by another vacancy. How this happens is the significant question. Similarly, the proposition that Al atoms can hop to metal vacancies on second nearest neighbor sites, or even more distant sites, does not suffice to explain data regarding the interdiffusion of heterojunctions. If we take the case that AlAs is under a layer of GaAs, a few Al atoms might percolate into the GaAs by hopping into the vacancies there, but this would soon deplete all the vacancies in the GaAs and leave a corresponding excess of vacancies in the AlAs. It must also be possible to recycle the vacancies so that a plausible concentration of them can account for the interdiffusion of many percent concentrations of Al.

It may also be worth noting that a conclusion similar to that just given would occur if we thought we were dealing with host interstitial mediated diffusion. However, we are confident that host interstitials play very little role. This follows from, among other considerations,[2] differential thermal analysis[25] measurements of the energy release when a radiation induced Ga interstitial annihilates a Ga vacancy in GaAs is 8.5 eV while positron annihilation experiments show that all samples of GaAs contain single vacancies and vacancy complexes at concentrations[26] exceeding $1x10^{16}$ cm^{-3}. Simple application of the Law of Mass Action that shows interstitial concentrations are less than one per cm^{-3}. The observation of As antisites in EL2 shows that there can be no significant concentration of As interstitials either.[27,28]

Let us begin with the case of Te doping, Eq.(6), for which our simulation indicates the simplest array of complexes and that the vacancies are migrating as single vacancies. We also provide in Table 7 the values of the enthalpies associated with the formation of the most prevalent ionization state of some relevant defects, which we have previously deduced[2,27] by semi-empirical means, for the cases that the Fermi level, E_f be: at the edge of the conduc-

tion band density of states distribution, E_c; at the intrinsic level, Ei_i; and at the edge of the valence band density of states distribution, E_v. Assuming that there are no divacancy or larger vacancy clusters, the Al diffusion may occur in n-type material via: a) 2nn hopping of V_{Ga}; b) 1nn hopping of V_{Ga}; or c) 1nn hopping of V_{As}. To avoid leaving behind a trail of antisite defects, mechanisms b) and c) require an 11-step process which has at its saddle point a ring of five antisites plus the opposite vacancy, i.e., for V_{Ga}, the string $V_{As}As_{Ga}Ga_{As}As_{Ga}Ga_{As}As_{Ga}$ running around one six-fold ring of the zinc-blende structure. If we take the enthalpy of formation of this configuration to be twice that of the antistructure pair plus that for the one isolated antisite plus that of the original V_{Ga}^- then we get 3.8 eV. Adding to this the 1nn hopping enthalpy of the As atom that must hop next (with no correction for its bonding), we get 4.8 eV as the estimate for the activation energy long ranged migration of Al, Q(Al), by mechanism b). This may be something of an overestimate because a large cluster of antisites costs less than the sum of the components; indeed, the inversion of the entire crystal could be done for zero net enthalpy of formation. However, it seems unlikely that this overestimate could bring the value for mechanism b down to the reported value, 2.9 eV. Similarly, we estimate Q(Al) enthalpy via mechanism c) to be 5.0 eV, which is even less viable. However, for mechanism a) we estimate Q(Al) to be the sum of 0.8 eV to make the ionized V_{Ga} and 2.4 eV for the 2nn hop of the Ga atoms which must occur for long ranged diffusion of Al to proceed, i.e., to recycle the vacancies and avoid the objection to percolation proposals (see Fig. 1). This value, 3.2 eV is 10% greater than that reported, which we claim to be good agreement. The linear variation of D(Al) with n is clearly explained by the single acceptor ionization level of V_{Ga}, which is deduced also from many other calculations and observations. (See Ref. 27 for a review.) Mechanism

a) would cause atomic mixing only on the cation sublattice, but one might expect a similar mechanism with V_{As} and anion 2nn hopping to occur and to mix the anion sublattice.

The difficult point is now to explain why the same V_{Ga} 2nn hopping mechanism a) does not dominate in intrinsic GaAs and lead to Q(Ga) = 4.0 eV, because the only difference would be the greater enthalpy to form the V_{Ga}. We regard the following explanation as speculative. We note that the pre-exponential factor that goes with the Q(Ga) for i-GaAs is very large,[23] 2.9×10^8 cm^2/s, compared to values of order 1 that are typical of simple mechanisms, such as a). Such values are usually explained in terms of contributions to the entropy of the process due to ionization of the defects; the effect of REM, recombination enhancement with the thermal equilibrium concentrations of electrons and holes that recombine, and form, at the migrating and ionized defects has also been noted recently.[29] Carrier recombination would be much more expected to enhance the 1nn ring mechanisms b) and c) than the simple single V_{Ga} 2nn hop mechanism a). Therefore, we propose that a) is simply swamped out by b) and c) under intrinsic conditions because, although they have larger Q(Ga)'s, they also have much larger pre-exponential factors. It is noted in Ref. 29 that the net Q for these thermal carrier REM processes contains a contribution for the enthalpy to form the carriers that recombine, which is approximately the enthalpy of the band gap, about 1.5 eV for GaAs. Thus if we add 1.5 eV for the carriers to the 4.8 eV estimated above for the 1nn ring mechanism b) we have 6.3 eV. As both estimates are thought to be overestimates, the agreement with the reported value, 6 eV, is quite good. Mechanisms b) and c) both imply substantially equal degrees of atomic mixing on both sublattices.

The question of the enhancement of Al diffusion upon adding Zn was first treated in Ref. 5 in terms of mechanism c). It is noted that the effect of Zn on D(Al) is much more dramatic than that for any other impurity enhanced migration effect in AlGaAs. It is a factor of 10^{10} or more. It seems unlikely that such a large factor could occur by any effect to the pre-exponential. Thus, an explanation in terms of the effect of Zn directly on the Q was sought. Recall that Zn diffuses into GaAs as an interstitial donor and occupies vacant Ga sites to cause p-type doping. As Zn_i diffusion vastly faster than any vacancies, the initial effect of introducing Zn must be to wipe out most V_{Ga} and thus greatly to suppress mechanisms a) and b). It would also seem to covert any divacancies or larger clusters present to V_{As}. When there is time to re-equilibrate the vacancy populations, the effect of p-type doping is also to generate more V_{As} in their single positive ionization state. Thus, we come to V_{As} 1nn hopping ring mechanism c). In Ref. 5 it was argued that the interaction between the oppositely charged Zn interstitial and the V_{Ga}Ga/As multiple acceptor complex that forms when V_{As} hops to a nearest neighbor site could reduce the Q for process c) by as much as 3 eV. The empirical value might appear to be smaller still by the enthalpy to form the V_{As} if the experiment is performed on a short time scale so that vacancies do not equilibrate with the surface but are supplied by the vacancy populations of the initial sample. (For this, the effect of Zn in breaking divacancy and larger clusters would be important.) Thus, we can argue that the Q(Al) for c) could be 4.0 eV less than obtained by adding the estimated values from Table 7, which is 6.4 - 4.0 = 2.4 eV versus 1 eV. This also implies essentially equal mixing of both sublattices, which is sometimes found[6] and sometimes not found.[4]

At this writing we can only speculate that such large vacancy clusters have formed during the preparation of the samples for which atomic mixing is largely on the cation sublattice, that trivacancy (or larger clusters) form even after Zn is introduced. Again we must assume the enthalpy to form these vacancies does not appear in the empirical Q(Al) because there is no time to equilibrate with the surface. Then we consider the vacancy cluster mechanism attributed to Si doping just below. The finding of an n^{+3} variation of D(Al) for Si-doping strongly implies that the mediating defect complex is a triple acceptor. This is an effect on the pre-exponential factor, most probably on the number of defects present to mediate diffusion, and not on the Q. One might suppose that this complex would be the divacancy, which is likely a single acceptor, plus a metal antisite, which is a double acceptor.However, the VIDSIM simulation keeps presenting us with a trivacancy plus metal antisite as the prevalent, rapidly diffusing, Al entraining complex. This should also be a triple acceptor complex.

Consider the enthalpy of formation of the trivacancy - Al antisite complex. In Table 7 we find it should cost 2.1 eV to make a divacancy plus 0.8 eV to make a V_{Ga}. When we let the V_{Ga} bind to the V_{As} of the divacancy, we have assumed for VIDSIM that 1 eV is released.Thus, we estimate the trivacancy to cost only 1.9 eV. One may worry that this is less than the cost of the divacancy, so let us keep in mind that the true value might be somewhat greater. We take the enthalpy of formation of the Al antisite to be the same as the Ga, 0.35 eV, so we have a tentative sum of 2.25 eV.

Consider in Fig. 2 how the trivacancy - metal antisite complex migrates. Of course, the antisite can run down the length of the trivacancy to make the net reaction

$$(V_{Ga}V_{As}V_{Ga}Al_{As})^{-3} \text{ <==> } (Al_{Ga}V_{As}V_{Ga}V_{As})^{0} + 3e^{-} , \quad (10)$$

but this does not lead to any long ranged migration. (We do discuss the interest of this reaction with possible regard to DX below.) In order for a complex involving any V_{As}'s to migrate significant distance, it is necessary for atoms on the As sublattice to move. Thus we suppose as the first step that one of the three As atoms at the end of the chain away from the Al hops to the V_{Ga} there. Our Ballistic model estimate of the activation enthalpy for that process is 1.0 eV. As noted our extensive MIS measurements support that model estimate and we find it mostly likely that hopping As atom does not know what it will encounter after it reaches that first step site. Thus, the net activation barrier to this point is tentatively estimated at only 3.25 eV and is suspected to be something of an underestimate. Once the As reaches step 2 in Fig. 2, it has vacancies on two sides and its other to bonds are to As rather than to Ga or Al. Therefore, it will hop again almost immediately and with very little activation barrier. Assuming it goes to step 3 and not back to 1, it finds itself with two good bonds but still between two vacancies. Thus, it will again have a smaller activation barrier to hop again than it did at step 1. if it goes to step 4 rather than back to 2, it will find itself bonded to two As, one Al and one vacancy - the best home it has had since it left 1. At this point, we suppose the Al antisite 2nn hops to one of the two V_{As} available, step 5, for a cost of somewhat less than 0.9 eV that would obtain it were on its proper lattice site. Adding 0.9 to the initial 2.25 gives 3.15 eV to which should be added some part of the 1.0 eV barrier when the As hopped. It seems our tentative estimate is very close to the reported 4 eV. Finally in step 6 the As regains its sublattice on the other side of the chain with almost no activation barrier.

In Fig. 3, we show how the trivacancy complex might migrate further.

Having proposed an explanation for the n^3 dependence and the Q = 4 eV, we must explain why the mixing is 300 times greater on the metal sublattice than on the anion sublattice. We note that each As atom is moved only across one six-fold ring as a result of this process. On the other hand, the Al antisite remains entrained with the trivacancy and thus moves a much longer distance before separating from the trivacancy. Because of the square root of number of steps variation of a random walk process, one gets further with long steps than with proportionately more short steps. Finally we return to reaction (10) by which the trivacancy complex transforms between a neutral donor (in n-type) and a triple acceptor. This is the property of the DX center in AlGaAs that has previously been attributed to the single As vacancy in some models.[30] However, previously unexplained facts regrading DX are that the extrapolation of its ionization level with Al concentration to pure GaAs fails to reach the value found in pure GaAs rather badly[31] (260 meV above E_c versus 140 meV) and the relative strength of the fine structure in the DLTS spectrum[32] generally attributed to alloy disorder are not in the simple relation given for a random distribution of Al. (Also the high temperature side of the spectrum is not as expected.) If we suppose that the X of the DX is the trivacancy rather than a single vacancy, then the non-random distribution of Al with these complexes may resolve these quandaries.

SUMMARY

We have introduced VIDSIM, a computer simulation program that allows one to run very large simulations on widely available personal computers at almost no cost by running them in background while the machines are being used for word processing etc. or would otherwise stand idle. VIDSIM allows one to work

out the complex macroscopic consequences of his estimates of atomic scale parameters through many millions of defect migrations at interactions. We find that our own assumptions lead to consequences, particularly the formation of trivacancies and their entrainment of Al, that we did not expect. Ref. 18 shows that the assumptions of others regarding the consequences of the so-called kick out model for diffusion of Au into Si are not correct. We have given a rather preliminary account of how previously published estimates of atomic scale thermochemical parameters, Refs. 2 and 28, might reconcile an array of empirical reports regarding interdiffusion across AlAs/GaAs interfaces that must initially seem quite confusing. We have also noted that the simulation gives detailed accounts of the point defect complex near the interface that result from the assumptions made regarding the interdiffusion. These may be useful in estimating the consequent effects on the transport properties of the junctions. Finally, we have noted a possible elucidation of two of the mysteries regarding the DX complex.

This work was supported in part by the Air Force Office of Scientific Research AFOSR-86-0309.

REFERENCES

* Present address: Max Planck Institut fur Festkorperforschung, Stuttgart, Federal Republic of Germany.

1. F.A. Kroger, *Chemistry of Imperfect Crystals* (North Holland, Amsterdam, 1964).
2. J.A. Van Vechten in *Handbook on Semiconductors Vol. 3 Materials, Properties and Preparation* edited by S.P. Keller (North Holland, Amsterdam, 1980) chapter 1.
3. P. Mei, S.A. Schwarz, T. Venkatesan, C.L. Schwartz, J.P. Harbison, L. Florez, N.D. Theodore, and C.B. Carter, *Appl. Phys. Lett. 53*, 2650 (1988).
4. A recent review of the literature on this point is given by D.G. Deppe and N. Holonyak, *J. Appl. Phys. 64*, R93 (1988).
5. J.A. Van Vechten, *J. Appl. Phys. 53*, 708 (1982).
6. M.D. Camras, N. Holonyak, K. Hess, M.J. Ludowise, W.T. Dietz, and C.R. Lewis, *Appl. Phys. Lett. 42*, 185 (1983).
7. S.A. Schwarz, P. Mei, T. Venkatesan, R. Bhat, D.M. Hwang, C.L. Schwarz, M. Koza, L. Nazar, and B.J. Skromme, *Appl. Phys. Lett. 53*, 1051 (1988).
8. D.G. Deppe, W.E. Plano, J.E. Baker, N. Holonyak, M.J. Ludowise, C.P. Kuo, R.M. Fletcher, T. Osentowski and M.G. Craford, *Appl. Phys. Lett. 53*, 2211 (1988).
9. J.A. Van Vechten, *Phys. Rev. B 12*, 1247 (1975).
10. J.A. Van Vechten and J.F. Wager, *Phys. Rev. B 32*, 5259 (1985).
11. M.T. Juang, J.F. Wager, and J.A. Van Vechten, *J. Electrochem. Soc. 135*, 2019 and 2023 (1988).
12. J.F. Wager and J.A. Van Vechten, *Phys. Rev. B 32*, 5251 (1985).

13. M.J.L. Sangster and A.M. Stoneham, *J. Phys. C 17*, 6093 (1984).

14. J.A. Van Vechten, to be published.

15. R.W. Keyes, *J. Chem. Phys. 29*, 467 (1958).

16. R.M. Emrick, *Phys. Rev. 122* (1961); *Phys. Rev. B 22*, 3563 (1980).

17. G.A. Samara, *Phys. Rev. B 37*, 8523 (1987).

18. U. Schmid, Thesis, Oregon State University, 11/17/1988.

19. J.W. Corbett and G.D. Watkins, *Phys. Rev. 138*, A555 (1965).

20. L.J. Cheng, J. Corelli, J.W. Corbett, and G.D. Watkins, *Phys. Rev. 152*, 761 (1966).

21. C.A.J. Ammerlaan and G.D. Watkins, *Phys. Rev. B 5*, 3988 (1972).

22. J.A. Van Vechten, *Phys. Rev. B 33*, 2674 (1986).

23. T.Y. Tan and U. Goesele, *Appl. Phys. Lett. 52*, 1240 (1988).

24. M.E. Greiner and J.F. Gibbons, *Appl. Phys. Lett. 44*, 750 (1984).

25. H.J. Lim, H.J. von Bardeleben, and J.C. Bourgoin, *J. Appl. Phys. 62*, 2738 (1987).

26. S. Dannefaer and D. Kerr, *J. Appl. Phys. 60*, 591 (1986).

27. J.F. Wager and J.A. Van Vechten, *Phys. Rev. B 35*, 2330 (1987).

28. J.F. Wager and J.A. Van Vechten, *Phys. Rev. B 39*, 1967 (1989).

29. J.A. Van Vechten, *Phys. Rev. B 38*, 9913 (1988).

30. J.A. Van Vechten, *Mat. Res. Soc. Symp. Proc. 46*, 83 (1985).

31. T.N. Theis, P.M. Mooney, and S.L. Wright, *Phys. Rev. Lett. 60*, 361 (1988).

32. P.M. Mooney, T.N. Theis, and S.L. Wright, *Appl. Phys. Lett. 53*, 2546 (1988).

Table 1

Nearest neighbor hopping "kinetic energy" values, ΔH_k^{Inn}(A: GaAs), ascribed to various atoms and, as explained in text to a vacancy, in GaAs used in present simulation. (Values in eV.)

A = Vac	Al	Si	Ga	As	Te
1.0	0.37	0.39	0.95	1.05	1.62

Table 2

Estimated atomic bond energies, E_j, for GaAs used in present simulation. (Values in eV.)

Al–Vac	Si–Vac	Ga–Vac	As–Vac	Te–Vac	Al–Al	Al–Si	Al–Ga
-1.0	-1.0	-1.0	-1.0	-1.0	-0.06	0.04	-0.0576

Al–As	Al–Te	Si–Si	Si–Ga	Si–As	Si–Te	Ga–As	Ga–Ga
0.08	0.02	0.12	0.03	0.04	0.01	0.06	-0.0576

Ga–Te	As–As	As–Te	Te–Te
0.02	-0.0625	-0.08	-0.1

Table 3

Histogram of configurations of vacancies found when initially 4 As vacancies are introduced into the 4% AlGaAs simulation. Sum over 22 T = 1100 K runs after sample is quite well annealed. "Snap-shot" state is denoted as SS; state after next hop is denoted N. Vacancies on As site are denoted V_b; vacancies on the metal site are denoted V_a. $V_aV_bV_a$ is a nearest neighbor tri- $V_aV_b + V_a$ is a divacancy plus a single (generally an antisite separates them). $3V_a$ is an antisite surrounded by three metal vacancies.

	V_a	V_b	$3V_a$	$3V_b$	$V_aV_bV_a$	$V_bV_aV_b$	$V_aV_b+V_a$	$V_aV_b+V_b$	ΣV_a	ΣV_b
SS	12	10	6	1	9	1	2	3	57	31
N	7	15	2	0	9	7	1	1	45	43

Table 4

Numbers of various types of antisite complexes found after arbitrarily chosen T = 1100 runs starting with 4 As vacancies.

	As_a	Al_b	Ga_b	As_aAl_b	As_aGa_b	$As_aGa_bAs_a$	$Ga_aAs_bGa_a$
Run 41	24	2	23	1	6	1	0
Run 42	21	2	24	1	9	1	0
Run 52	27	3	26	1	7	1	0
Run 61	28	3	27	1	4	2	1
Run 71	27	3	28	1	3	0	0

	$As_aAl_bAs_a$	$Al_bAs_aGa_b$	$Ga_bAs_aGa_bAs_a$	$Ga_bAs_aGa_bAs_aGa_b$
Run 41	0	0	0	0
Run 42	0	0	0	1
Run 52	0	0	0	1
Run 61	0	1	0	0
Run 71	1	1	1	1

Table 5

Histogram of configurations found after 4 metal vacancies are introduced in the 4% AlGaAs simulation. Sum over 22 T = 1100 K runs. Notation as in Table 3. Unlike Table 3, we here are able to specify the configuration of the Al nearest neighbors because they are so few and so simple.

	V_a	V_b	V_aAl_b	Al_aV_b	$Al_aAl_bV_a$	$Al_aV_bAl_a$
SS	50	34	2	1	1	0
N	50	34	1	2	0	1

Table 6

Number of various types of antisite complexes found after annealing and 22 T = 1100 K runs when starting from 4 metal vacancies and Te doping.

As_a	Al_b	Ga_b	As_aGa_b	$As_aAl_bAs_a$	V_aGa_b	As_aV_b	V_a
12	2	10	6	1	1	2	1

Table 7

Enthalpies associated with formation of ionized defects deduced in Refs. 2 and 26 by semi-empirical means. Values are in eV.

$E_f = E_c$		$E_f = E_i$		$E_f = E_v$	
V_{Ga}^{-}	0.8	V_{Ga}^{-}	1.6	V_{Ga}^{0}	2.3
V_{As}^{0}	2.3	V_{As}^{+}	1.8	V_{As}^{+}	1.0
As_{Ga}^{0}	1.6	As_{Ga}^{0}	1.6	As_{Ga}^{2+}	0.35
Ga_{As}^{2-}	0.35	Ga_{As}^{2-}	1.9	Ga_{As}^{0}	3.1
$V_{Ga}V_{As}^{-}$	2.1	$V_{Ga}V_{As}^{0}$	1.9	$V_{Ga}V_{As}^{+}$	2.3
$As_{Ga}Ga_{As}^{0}$	0.7	$As_{Ga}Ga_{As}^{0}$	0.7	$As_{Ga}Ga_{As}^{0}$	0.7

FIGURE CAPTIONS

Fig. 1 Proposed mechanism of long ranged Al diffusion for n-type, Te-doped AlAs/GaAs D(Al) = const. $n^{+1}\exp(-2.9\ eV/kT)$ is reported. Note that it is not enough for the Al to 2nn hop; Ga must also 2nn hop. It is the larger hopping energy of the Ga which determines the net activation barrier.

Fig. 2 Trivacancy - Al complex proposed mechanism for n-type AlAs/GaAs system for D(Al) = const. $n^{+3}\exp(-4\ eV/kT)$.

Fig. 3 Trivacancy - Al complex diffuses further.

Fig. 4 Trivacancy - Al complex converts between its triple acceptor and neutral donor configurations. We propose this may be the X of the DX center and note its non-random variation with Al concentration.

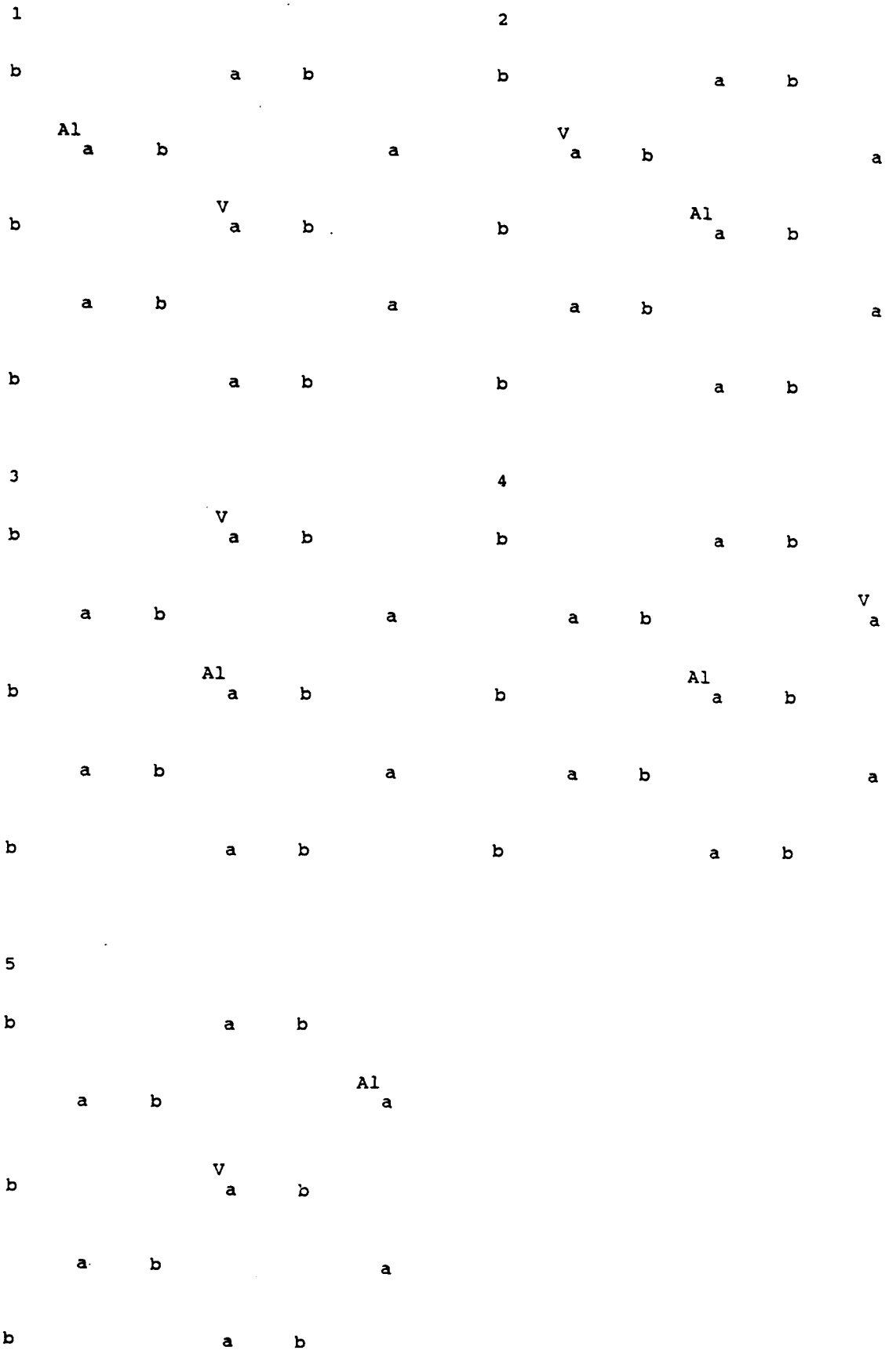

1
b a b
Al
a b a
V
b a b
a b a
b a b
2
b a b
V
a b a
Al
b a b
a b a
b a b
3
V
b a b
a b a
Al
b a b
a b a
b a b
4
b a b
V
a b a
Al
b a b
a b a
b a b
5
b a b
Al
a b a
V
b a b
a b a
b a b

FIG. 1

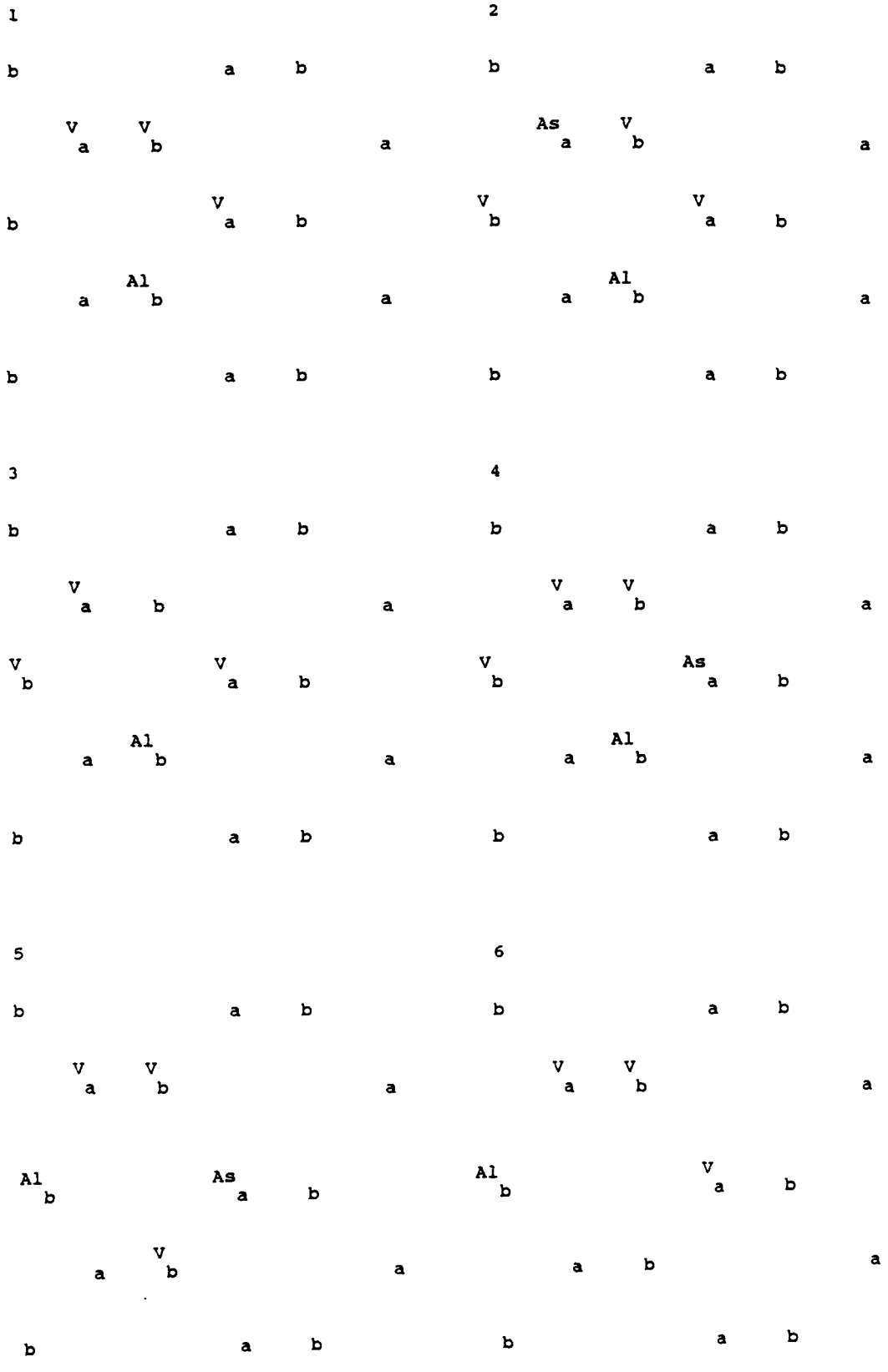

1
b a b
V V
a b
a
V
b a b
Al
a b
a
b a b
2
b a b
As V
a b
a
V
b a b
Al
a b
a
b a b
3
b a b
V
a b
a
V V
b a b
Al
a b
a
b a b
4
b a b
V V
a b
a
V As
b a b
Al
a b
a
b a b
5
b a b
V V
a b
a
Al As
b a b
V
a b
a
b a b
6
b a b
V V
a b
a
Al V
b a b
a b
a
b a b

Fig 2.

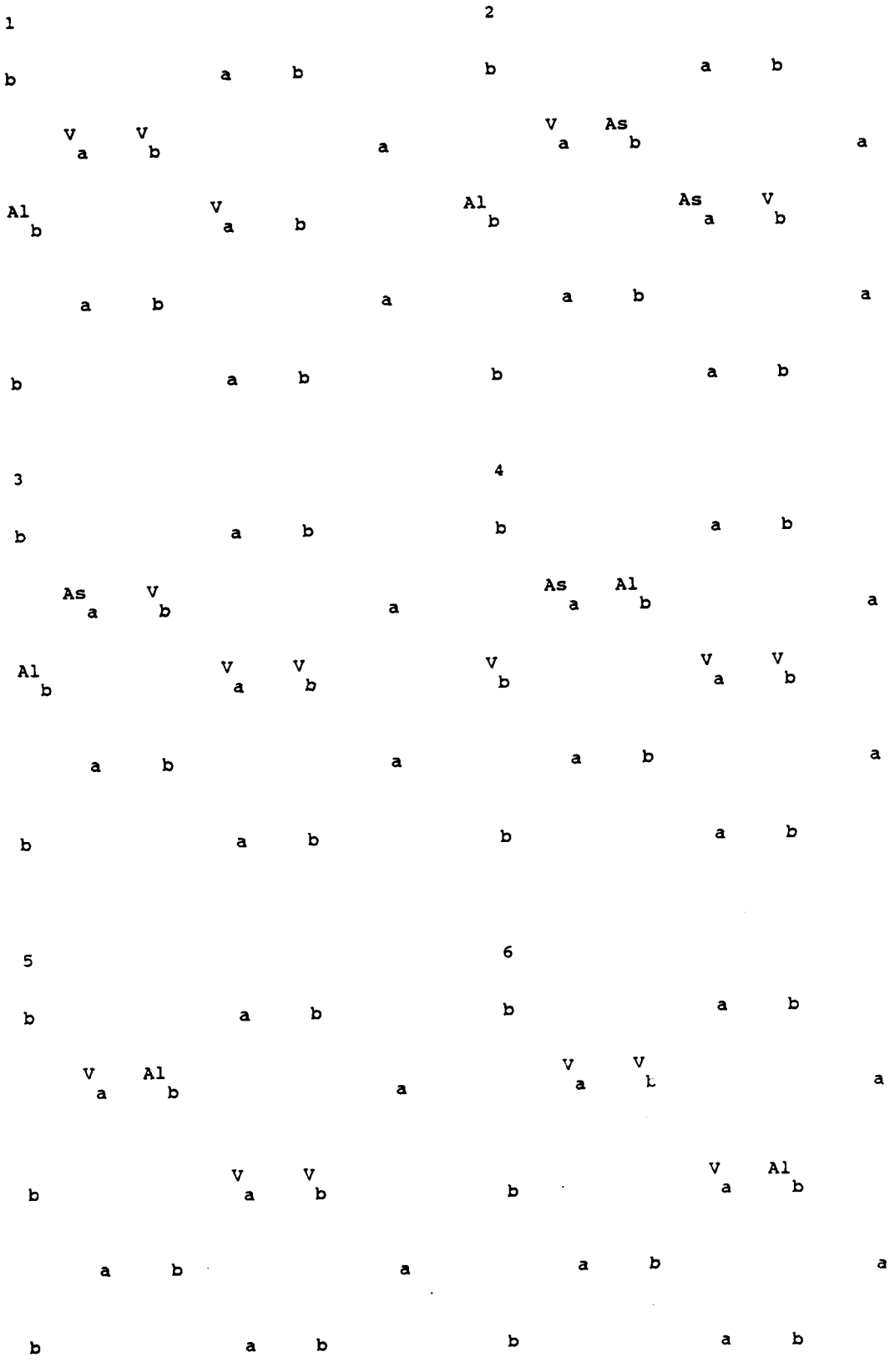


FIG 3


1
b a b
V V
a b a
Al V
b a b
a b a
b a b
2
b a b
V Al
a b a
V V
b a b
a b a
b a b
3
b a b
V V
a b a
V Al
b a b
a b a
b a b

Fig. 4

ENTROPY OF MIGRATION FOR ATOMIC HOPPING

T.W. Dobson, J.F. Wager, and J.A. Van Vechten

Department of Electrical and Computer Engineering

Center for Advanced Materials Research

Oregon State University

Corvallis, Oregon 97331

ABSTRACT

The entropy of migration of an atom undergoing an atomic hop into a nearest neighbor vacancy is calculated assuming the entropy to be translational at the saddle point configuration and vibrational at the initial, equilibrium state prior to the hop. These contributions to the entropy of migration are calculated using standard statistical thermodynamic expressions, assuming that the translational entropy is given by the particle in a box approximation and the vibrational entropy is that of a simple harmonic oscillator vibrating at the Debye frequency. This formulation of the entropy of migration is then quantitatively applied to explain the abnormal prefactors experimentally deduced in InP drain current drift measurements, deep-level defect transformation kinetic studies of the metastable M center in InP, and in Si and Ge self-diffusion experiments.

I. INTRODUCTION

The kinetics of solid state processes such as diffusion and defect transformation reactions are often found to exhibit the Arrhenius form,[1]

$$k = A \exp(-E_a/k_B T) \tag{1}$$

where k is the rate constant, A is the prefactor, E_a is the activation energy, and k_B is Boltzmann's constant. Adopting the thermodynamic formulation of rates,[1] it is possible to make the following identifications,

$$A = A_o \exp(\Delta S_a/k_B) \tag{2}$$

$$E_a = \Delta H_a \tag{3}$$

where A_o can loosely be interpreted as an attempt frequency and ΔS_a and ΔH_a are, respectively, the entropy and enthalpy of activation. These activation quantities represent the difference of the reactant at its saddle point configuration with respect to its initial configuration.

The prefactor, A, has often been observed experimentally to be abnormally large, of alternatively, abnormally small, in a variety of solid state processes such as self-diffusion[2-7] and deep-level defect transformations.[8-14] The unusual magnitude of the prefactor is attributed to the correspondingly abnormal magnitude of the entropy of activation. The physical reason for the unusual magnitudes of these empirically determined activation entropies is presently unclear.

Three contributions to the entropy of activation have been discussed in the literature:

1. A vibrational contribution due to a modification of the vibrational frequencies of the lattice when the reactant achieves its saddle point configuration.[15,16]
2. A configurational entropy due to the different multiplicity of bonding configurations (e.g. Jahn-Teller distortion) at the saddle point compared to the initial state.[17]
3. The entropy of ionization due to the softening of the lattice when an electron or hole is emitted from a defect state compared to when it is localized at that state.[16,18]

Note that ionization must occur concomitant with the process of interest for mechanism 3 to be operative. These physical mechanisms are of insufficient magnitude to explain the abnormal activation entropies often found experimentally.

The purpose of the work discussed herein is to propose a physical mechanism associated with atomic migration which is capable of explaining the very large magnitudes of the activation entropies. The proposal is motivated by the ballistic model (BM)[16,19,20] hypothesis that atomic migration into a vacancy on a nearest-neighbor site may be calculated by assuming that the migration energy is kinetic in origin. Within the context of this model, the migrating atom must attain a sufficient kinetic energy to successfully traverse the saddle point configuration. In the spirit of the BM, we propose that the activation entropy of a process involving atomic migration (e.g. diffusion, metastable defect transformations) will be determined (at least to a large extent) by the difference of the translational entropy of the hopping atom in its saddle point configuration and the vibrational entropy of the hopping atom in its equilibrium position prior to the hop. These two

contributions to the activation entropy are calculated using standard statistical thermodynamic expressions.

Using this formulation for the entropy of activation of atomic migration, we provide quantitative justifications for experimentally obtained prefactors for the activation entropy for phosphorous vacancy nearest-neighbor hopping contributions to drain current drift in InP MIS capacitors and field-effect transistors, for the prefactors found in defect transformation experiments involving the M center in InP and for the self-diffusion of Si and Ge. These explanations provide strong support for the BM and for our kinetic energy formulation of the activation entropy.

II. ENTROPY OF MIGRATION FOR NEAREST-NEIGHBOR HOPPING

The physical picture adopted in the BM is as follows. At temperatures above the Debye temperature, θ_D, the thermal vibrations of the atoms surrounding a vacancy fluctuate with a characteristic frequency given approximately by the Debye frequency, $\nu_D = k_B\theta_D/h$ where h is Planck's constant. Periodically these thermal fluctuations conspire to provide a path for an atom to hop into the vacancy at a cost of very little potential energy. The venue for this path is limited by the frequency of the thermal vibrations. For an atom to successfully migrate it must make the hop in a time less than the lifetime of the favorable venue which requires a minimum velocity of the hopping atom to be

$$v = \nu_D d \tag{4}$$

where d is the distance between lattice sites. This velocity corresponds to a kinetic energy

$$KE = \frac{1}{2} mv^2 \tag{5}$$

where m is the mass of the hopping atom. Thus, since the potential energy that the atom must overcome is assumed to be quite small, the entropy of migration is given by

$$\Delta H_m = \frac{1}{2} mv^2 = \frac{1}{2} m(Fd\ \nu_D)^2 \tag{6}$$

where F is a geometric constant equal to 0.9 in the case of a diamond or zincblende lattice.[19]

In order to calculate the entropy of migration, we extend the BM treatment. At temperatures above the Debye temperature an atom on a normal lattice site can be modeled as a simple harmonic oscillator vibrating at the Debye frequency. The entropy associated with this vibrating atom can be calculated from statistical thermodynamics and is given by,[21]

$$\frac{S_v}{k_B} = 3 \left\{ \frac{h\nu_D/k_BT}{\exp(h\nu_D/k_BT)-1} - \ln[1 - \exp(-h\nu_D/k_BT] \right\} \tag{7}$$

where the factor 3 is due to the three degrees of freedom of the oscillating atom.

According to the BM, during the limit venue that the atomic migration actually occurs, i.e. when an atom moves through the saddle point configuration, the hopping atom is essentially a free particle moving ballistically. Thus, for this brief period of time the vibrational modes are replaced by a translational mode. We can calculate the translational entropy using the particle in a box approximation,[21]

$$\frac{S_t}{k_B} = \ln \left[\frac{(2\pi emk_BT)^{3/2}}{h^3} V \right] \tag{8}$$

where e = 2.718 and V is the volume of the box which we take to be the volume occupied by two nearest-neighbor atoms. The entropy of migration of an atom hopping into a vacancy is thus given by,

$$\Delta S_m = S_t - S_v \quad . \tag{9}$$

Therefore, our formulation for the entropy of atomic migration is given by Eqs. (7)-(9). We will now employ these equations to explain the abnormal activation entropies found experimentally in self-diffusion and defect transformation reactions.

III. ENTROPY OF MIGRATION EXAMPLES

A. Phosphorous Vacancy Nearest-Neighbor Hopping

Phosphorous vacancy nearest-neighbor hopping (PVNNH) has been linked[13,14] to drain current drift (DCD) in InP metal-insulator-semiconductor-field-effect transistors (MISFETs). Consider an InP MISFET or MIS capacitor which has P vacancies in the channel region under its gate. Nearest-neighbor hopping of an In atom into the P vacancy can be described by the following defect reaction,

$$In_{In} + V_P^+ + 4e^- = V_{In}^- \, In_P^{-2} \quad . \tag{10}$$

Application of the Law of Mass Action to this defect reaction results in

$$\frac{[V_{In}^- \, In_p^{-2}]}{[V_P^+][In_{In}]} \alpha \, [e^-]^4 \quad . \tag{11}$$

The application of a positive gate bias to an InP MIS capacitor or MISFET induces an accumulation of electrons in the channel and hence an increase in $[e^-]$ which shifts the equilibrium to the right-hand-side of the defect

reaction. Four of the accumulated channel electrons are captured for every P vacancy annihilated. DCD in MISFETs or flatband voltage shift in MIS capacitors is attributed to the loss of these electrons from the channel.

From the BM activation enthalpy of $\Delta H_m(V_P) = 1.2$ eV was predicted[13] which was found to be in good agreement with experimental values[14] deduced from variable-temperature bias-stress measurements of InP MIS capacitors. Additionally, a computer simulation of the effect of the PVNNH mechanism on flatband voltage shift vs. bias stress measurements was performed.[14] In the kinetic analysis it was found that the rate-limiting step in the total reaction given by Eq. (10) was,

$$V_P^o + e^- = (V_{In}In_P)^- \quad . \tag{12}$$

and an entropy of activation, ΔS_a, was used as an adjustable parameter (the only adjustable parameter) in the computer simulation to fit the experimental data. A value $\Delta S_a(\text{expt.})/k_B = 15.3$ was deduced from the computer simulation[14] compared to that originally estimated[13] by Van Vechten and Wager from

$$\Delta S_a/k_B = \ell n(8D_o/a^2\nu_D) \tag{13}$$

where D_o is the prefactor of the diffusion constant for In self-diffusion in InP. Using $D_o = 1\text{x}10^5$ cm/s as reported[3] by Goldstein gives $\Delta S_a/k_B = 17.5$.

To calculate the entropy of activation we recognize that the rate-limiting reaction given by Eq. (12) involves both In hopping and ionization so that the activation entropy is given by,

$$\Delta S_a = \Delta S_m + \Delta S_i \tag{14}$$

where ΔS_i is the entropy of ionization. We calculate ΔS_i using[16,18,22]

$$\Delta S_i(T) - \Delta S_{CV}(T) = \frac{\alpha T(T+2\beta)}{(T+\beta)^2} \tag{15}$$

where ΔS_{CV} is the entropy of the bandgap (i.e. the entropy of formation of free electron-hole pairs) and for InP[22] $\alpha = 6.63\times10^{-4}$ eV/K and $\beta = 162$K. ΔS_m is calculated using Eqs. (7)-(9) which results in an activation entropy $\Delta S_a(\text{theory})/k_B = 8.2 + 6.8 = 15.0$ (the parameters used in this calculation are summarized in Table 1). 325K was used in the calculation because it is the average temperature at which the activation energy was extracted in the computer simulation. We regard the close agreement between experiment (15.3) and theory (15.0) as strong support for the importance of the translational and ionization entropies in determining the total activation entropy.

B. <u>The Metastable M Center</u>

The metastable M center[8-12] is an electron-irradiation-induced defect complex in InP which exhibits two distinct configurations denoted A and B. There is a reversible transformation between the configurations which exhibits the following kinetics[8-9] as deduced from Arrhenius plots:

$$A \rightarrow B \text{ Stage 1:} \quad k = 10^{18}\exp[-0.40 \text{ eV}/k_BT] \qquad T = 110\text{K} \tag{16}$$

$$A \rightarrow B \text{ Stage 2:} \quad k = 10^{11}\exp[-0.42 \text{ eV}/k_BT] \qquad T = 160\text{K} \tag{17}$$

$$B \rightarrow A \quad : \quad k = 10^{7}\exp[-0.24 \text{ eV}/k_BT] \qquad T = 140\text{K} \tag{18}$$

We have proposed[12] an atomic model for the M center. In terms of this atomic model, the reversible transformation, A-B, may be written as follows,

$$(V^-_{In}\ P^+_{In})^o + 3e^- \underset{P\ hop}{=} P^+_{In} + V^+_P + P^+_{In} \tag{19}$$

where the parentheses around the first term denotes a Coulombic attractive interaction between the individual point defects and P hop identifies P atomic migration to a nearest-neighbor vacancy as the transformation mechanism. We have discussed[12] the atomic model in some detail; it is our present objective to offer a quantitative explanation for the three widely divergent prefactors found in the kinetic equations.

First note that the reversible transformation between configurations as specified by Eq. (19) involves the capture or emission of three electrons as well as a P hop. We propose that the wide range of empirically deduced prefactors results from differences in the rate-limiting step in the defect complex transformation. We now individually treat each of the kinetic equations.

1. A → B Stage 1

This process occurs at a temperature of 110K with a prefactor $A_o(110, \text{expt.}) = 10^{18}$. We assume that the rate-limiting step is a process involving a P hop and emission of a single electron,

$$(V_{In}^{-} \; P_{In}^{+})^{o} \underset{P \text{ hop}}{\rightarrow} P_{In}^{+} + V_{P}^{o} + P_{In}^{o} + e^{-} \; . \tag{20}$$

For this rate-limiting defect reaction the prefactor can be written as

$$A_o = \nu_D \exp[(\Delta S_m + \Delta S_i)/k_B] \; . \tag{21}$$

The parameters used to calculate A_o are summarized in Table 1. The calculated prefactor $A(110, \text{theory}) = 1.2 \times 10^{18}$ is in excellent agreement with that found experimentally. Note the very large value of the activation entropy, $\Delta S_a/k_B = 12.2$, and that the translational entropy, $S_t/k_B = 8.0$, is a dominant contribution.

2. A → B Stage 2

This process occurs at 160K with a prefactor $A_o(160,expt.) = 10^{11}s^{-1}$. We assume that the rate-limiting step is due to the emission of a single electron,

$$(V_{In}^{-}P_{In}^{+})^{o} \rightarrow (V_{In}^{-}P_{In}^{+})^{+} + e^{-} \quad . \tag{22}$$

The prefactor of the rate-limiting step corresponds to electron emission and can be written,

$$A_o = v_{th}\ \sigma\ N_C\ \exp(\Delta S_i/k_B) \tag{23}$$

where v_{th} is the thermal velocity, σ is the capture cross section, and N_c is the conduction band effective density of states. Assuming $\sigma = 10^{-15}$ cm^2, which is a typical capture cross section, we obtain $A_o(160,\ theory) = 9.6x10^{11}$ s^{-1} (see Table 1) again in good agreement with experiment.

3. B → A

At a temperature of 140K the prefactor of the B → A transformation is fond to be $A_o(140,\ expt.) = 10^7\ s^{-1}$. In accordance[9] with Levinson et al. we assume that the reverse reaction occurs in one stage with the capture of three electrons as well as a P hop,

$$P_{In}^{+} + V_P^{+} + P_{In}^{+} + 3e^{-} \xrightarrow{\text{P hop}} (V_{In}^{-}P_{In}^{+})^{o}. \tag{24}$$

The prefactor of this process can be written as,

$$A_o = \nu_D \exp[(\Delta S_m - 3\Delta S_i)/k_B \tag{25}$$

where the negative sign before the ionization entropy arises because of electron capture and a concomitant hardening of the lattice. As summarized in Table 1, this results in a prefactor $A_o(140,\ theory) = 4.0x10^8\ s^{-1}$ in

reasonable agreement with that found experimentally. Note that in this case the large entropy of the three captured electrons dominates, yielding a negative activation entropy.

C. Self-Diffusion in Si and Ge

A further application of our proposed work for the entropy of atomic hopping involves analysis of the experimentally deduced prefactors for self-diffusion in Ge and Si. The prefactor for self-diffusion, D_o, for Ge over a temperature range of 766-928C was found to be,[2]

$$D_o(\text{Ge, expt.}) = 7.8 \text{ cm}^2/\text{s} \ . \tag{26}$$

The prefactor for Si self-diffusion has been reported[6] over a very wide range (1-9000 cm^2/s). Demond et al. have shown that the rate of self-diffusion in Si does not obey[6] a simple Arrhenius law. They have, however, interpreted this as due to the superposition of two activated processes. For the temperature range 830-1200C the dominant process has a prefactor that falls within the range,

$$D \text{ (Si, expt., low T)} = 0.2\text{-}20 \text{ cm}^2/\text{s} \ . \tag{27}$$

for temperatures greater than 1200C the dominant process has a prefactor of order

$$D_o(\text{Si, expt., high T}) \simeq 2000 \text{ cm}^2/\text{s} \ . \tag{28}$$

We assume that self-diffusion in Ge and low-temperature self-diffusion in Si proceeds via neutral monovacancy diffusion. The prefactor for self-diffusion is given by,[23,24]

$$D_o = \frac{1}{8} a^2 f \nu_D \exp[(\Delta S_m + \Delta S_f)/k_B] \tag{29}$$

where a is the lattice constant, f is the correlation factor (equal to 1/2 for self-diffusion by vacancy migration in diamond lattices[23]), and ΔS_f is the entropy of formation of a vacancy. ΔS_m is calculated using Eqs. (7)-(9). Note that ΔS_m is temperature-dependent so that the average temperature over the experimental range is used to calculate ΔS_m. The parameters used to calculate ΔS_m for Si and Ge are summarized in Table 2. The entropies of formation were taken[4,25] to be $\Delta S_f(Si)/k_B = 3.0$, $\Delta S_f(Ge)/k_B = 2.6$. Using Eq. (29), we find

$$D_o(Ge, \text{theory}) = 11 \text{ cm}^2/\text{s} \tag{30}$$

$$D_o(Si, \text{theory, low T}) = 24 \text{ cm}^2/\text{s} \tag{31}$$

which are in reasonable agreement with that found experimentally, particularly when it is realized that precise values for ΔS_f are not well established and there is a large spread in the experimental data in the case of Si self-diffusion. Note that the total activation entropy of the prefactor (i.e. $\Delta S = \Delta S_m + \Delta S_f$) is quite large for both Ge and Si, $\Delta S_a(Ge, \text{theory})/k_B = 8.9$ and $\Delta S_a(Si, \text{theory}) = 9.2$.

An alternative mode of vacancy migration that yields the same theoretical prefactor for low temperature Si self-diffusion is that singly ionized vacancies, rather than neutral vacancies, are the mediators of self-diffusion and that they are de-ionized through capture of free carriers concomitant with the hopping event, i.e.

$$V_{Si}^{+} + e^{-} \xrightarrow[\text{hop}]{} V_{Si}^{x} \tag{32}$$

or

$$V_{Si}^{-} + h^{+} \xrightarrow[\text{hop}]{} V_{Si}^{x} . \tag{33}$$

The activation entropy for both of these self-diffusion mechanisms is the same as in Eq. (29), namely $(\Delta S_m + \Delta S_f)/k_B$, since the ionization entropy contribution from the singly ionized vacancy is canceled by the ionization entropy associated with the free carrier capture. Note that Eqs. (32) and (33) for Si are similar to Eqs. (12), (20), and (24) for InP in that our prefactor analysis suggests that carrier capture or emission often occurs concomitant with atomic hopping.

The prefactor for high temperature self-diffusion in Si can be justified in accordance with our activation entropy formulation if the total activation entropy used in Eq. (29) is given by

$$\Delta S_a = \Delta S_m + \Delta S_f + \Delta S_i \ . \tag{35}$$

Using an average temperature of 1573K and Eq. (15) to evaluate ΔS_i with[16] $\alpha = 4.73 \times 10^{-4}$ eV/K and $\beta = 636$K yields, $\Delta S_i/k_B = 5.0$. This leads to (see Table 2), $\Delta S_a/k_B = 13.9$ and a prefactor for high temperature Si self-diffusion

$$D_o(\text{Si, theory, high T}) = 2708 \ \text{cm}^2/\text{s} \tag{36}$$

which is in reasonable agreement with that deduced by Demond et al. There are various mechanisms for self-diffusion which would result in this prefactor: neutral vacancies which ionize concomitant with the hop to become singly ionized, singly ionized vacancies which hop without changing their ionization state, or doubly ionized vacancies which undergo free carrier capture during the hop to become singly ionized. The situation becomes further complicated if the negative U character[26] of V_{Si} is brought into consideration.

IV. CONCLUSIONS

We propose the entropy of atomic migration may be calculated from a difference of the translational and vibrational entropy as calculated using Eqs. (7)-(9). It is found that in a kinetic process involving atomic hopping is often dominated by the effects of the migrational and ionization entropies. We have been able to quantitatively account for the prefactors, or equivalently, the activation entropies, of PVNNH in InP MIS devices, transformations in the metastable M center in InP and self-diffusion in Si and Ge. Prefactors with magnitudes as large as those exhibited in the processes discussed were previously considered to be anomalous. We believe that analysis of the magnitudes of prefactors, along the lines of that discussed herein, may be a powerful approach for the microscopic identification of defects and defect-related processes.

ACKNOWLEDGEMENTS

This work was supported by th Air Force Office of Scientific Research under Contract No. AFOSR 86-0309.

Table 1. InP parameters used in the phosphorous vacancy nearest-neighbor hopping and metastable M center calculations.

Lattice constant, a (Å)	5.86875
Debye temperature, θ_D (K)	292
Debye frequency, ν_D (s^{-1})	6.1×10^{12}
Mass, m (kg)	In = 1.91×10^{-25} P = 5.14×10^{-26}
Tetrahedral radius, r (Å)	In = 1.405 P = 1.128
Volume, V (m^{-3})	1.76×10^{-29}

Phosphorous Vacancy Nearest-Neighbor Hopping

Temperature, T (K)	325
Vibrational entropy, S_v/k_B	3.4
Translational entropy, S_t/k_B	11.6
Migrational entropy, $\Delta S_m/k_B$	8.2
Ionization entropy, $\Delta S_i/k_B$	6.8
Activation entropy, $\Delta S_a/k_B$	15.0

Metastable M Center

A → B Stage 1

Temperature, T (K)	110
Vibrational entropy, S_v/k_B	0.8
Translational entropy, S_t/k_B	8.0
Migrational entropy, $\Delta S_m/k_B$	7.2
Ionization entropy, $\Delta S_i/k_B$	5.0
Activation entropy, $\Delta S_a/k_B$	12.2
Prefactor, A_o (s^{-1})	1.2×10^{18}

Table 1. (continued)

A → B Stage 2	
Temperature, T (K)	160
Thermal velocity, v_{th} (cm/s)	3.1×10^{7}
Capture cross section, σ (cm^2)	10^{-15}
Density of states, N_C (cm^{-3})	1.04×10^{17}
Ionization entropy, $\Delta S_i/k_B$	5.7
Prefactor, A_o (s^{-1})	9.6×10^{11}
B → A	
Temperature, T (K)	140
Vibrational entropy, S_v/k_B	1.3
Translational entropy, S_t/k_B	8.4
Migrational entropy, $\Delta S_m/k_B$	7.1
Ionization entropy, $\Delta S_i/k_B$	-5.5
Activation entropy, $\Delta S_a/k_B$	-9.4
Prefactor, A_o (s^{-1})	5.0×10^{8}

Table 2. Parameters used to calculate the entropy of migration for Si and Ge self-diffusion. The values in parentheses refer to high temperature self-diffusion in Si.

Parameter	Silicon	Germanium
Lattice constant, a (Å)	5.43072	5.65754
Debye temperature, θ_D (K)	648	374
Debye frequency, ν_D (s^{-1})	1.35×10^{13}	7.79×10^{12}
Mass, m (kg)	4.66×10^{-26}	1.21×10^{-25}
Tetrahedral radius, r (Å)	1.173	1.225
Volume, V (m^{-3})	1.35×10^{-29}	1.54×10^{-29}
Temperature, T (K)	1288 (1573)	1120
Vibrational entropy, S_v/k_B	5.1 (5.7)	6.3
Translational entropy, S_t/k_B	11.3 (11.6)	12.6
Migrational entropy, $\Delta S_m/k_B$	6.2 (5.9)	6.3
Formation entropy, $\Delta S_f/k_B$	3.0	2.6
Ionization entropy, $\Delta S_i/k_B$	(5.0)	
Activation entropy, $\Delta S_a/k_B$	9.2 (13.9)	8.9
Prefactor, D_o (cm^2/s)	24 (2708)	11

REFERENCES

1. M. Boudart, Kinetics of Chemical Processes, (Prentice-Hall, Englewood Cliffs, New Jersey, 1968).
2. H. Letaw, W.M. Portnoy, and L. Slifkin, Phys. Rev. 102, 636 (1956).
3. B. Goldstein, Phys, Rev. 121, 1305 (1961).
4. M. Lannoo and G. Allan, Phys. Rev. B 25, 4089 (1982).
5. J.A. Van Vechten, Phys. Rev. B 33, 8785 (1986).
6. F.J. Demond, S. Kalbitzer, H. Mannsperger, and H. Damjantschitsch, Phys. Lett. 93A, 503 (1983).
7. A.F.W. Willoughby, Mat. Res. Soc. Symp. Proc. 14, 237 (1983).
8. J.L. Benton and M. Levinson, Mat. Res. Soc. Symp. Proc. 14, 95 (1983).
9. M. Levinson, J.L. Benton, and L.C. Kimerling, Phys. Rev. B 27, 6216 (1983).
10. M. Levinson, M. Stavola, J.L. Benton, and L.C. Kimerling, Phys. Rev. B 28, 5848 (1983).
11. M. Stavola, M. Levinson, J.L. Benton, and L.C. Kimerling, Phys. Rev. B 30, 832 (1984).
12. J.F. Wager and J.A. Van Vechten, Phys. Rev. B 32, 5251 (1985).
13. J.A. Van Vechten and J.F. Wager, J. Appl. Phys. 57, 1956 (1985).
14. M.T. Juang, J.F. Wager, and J.A. Van Vechten, J. Electrochem. Soc. 135 (1988).
15. C.P. Flynn, Point Defects and Diffusion, (Clarendon, Oxford, 1972).
16. J.A. Van Vechten in Handbook of Semiconductors, Vol. 3, Edited by S.P. Keller (North-Holland, Amsterdam, 1980), Chapt. 1.
17. R.A. Swalin, J. Phys. Chem. Solids 18, 290 (1961).
18. J.A. Van Vechten and C.D. Thurmond, Phys. Rev. B 14, 3539 (1976).

19. J.A. Van Vechten, Phys. Rev. B 12, 1247 (1975).

20. J.A. Van Vechten and J.F. Wager, Phys. Rev. B 32, 5259 (1985).

21. J.F. Lee, F.W. Sears, D.L. Turcott, Statistical Thermodynamics, 2nd Edition (Addison-Wesley, Reading, Massachusetts, 1973).

22. C.D. Thurmond, J. Electrochem. Soc. 122, 1133 (1975).

23. C. Zener, J. Appl. Phys. 22, 372 (1951).

24. A. Seeger and K.P. Chik, Phys. Stat. Sol. 38, 3148 (1967).

25. J.A. Van Vechten, Mat. Res. Soc. Symp. Proc. 46, 83 (1985).

26. J.A. Van Vechten, Phys. Rev. B 33, 2674 (1986).

Mössbauer Spectra and the DX Center Complex in AlGaAs

J. A. Van Vechten

Center for Advanced Materials Research

Department of Electrical and Computer Engineering

Oregon State University,

Corvallis Oregon 97331, USA

ICP# = 7680, 7155

Abstract. Mössbauer spectra for the "DX" deep level defect in AlGaAs alloys are just now being made available, e.g., by Gibart et al. (1988). The interpretation of these spectra has been somewhat clouded by the fact that Mössbauer specialists have generally not considered the antisite defect among the defect complexes present in III–V crystals near the Mössbauer isotope that affect the various components of the spectra. We note that the antisite defect is well established as a component of e.g., the EL2 deep level defect, and generally as the most prevalent native point defect in III–V's. Furthermore, the spectra of Gibart et al. are consistent with the model that the "X" of the DX defect complex is the metal on As site antisite defect plus Ga vacancy pair that results from As vacancy nearest neighbor hopping. In particular, they show that a major fraction of donors remain in an unperturbed state when DX compensates the sample.

The "DX" phenomena in AlGaAs alloys, and some other semiconductors, is a spontaneous compensation of all types of n-type doping at low temperatures that may temporarily be suppressed by absorption of sub-bandgap light (Lang et al. 1977, 1979; Lang 1985; Gibart et al. 1988). The transition between the stable, largely non-conducting state and the metastable, conducting state has been attributed by some to a "large lattice relaxation" (LLR) and by others to a mainly electronic relaxation. The origin of this low temperature compensation effect and the "persistent photoconductive transition" has been the subject of extensive investigation and controversy. The subject has practical import because it is widely believed that AlGaAs and similar semiconductors can compete successfully against Si only if they can be made to operate at temperatures (T) below the freeze-out of acceptors in Si.

The earliest hypothesis (Lang et al., 1979) was that the donors, D, tend to form complexes with a native point defect, X, which was proposed to be the As vacancy, V_{As} = X, and that X had two configurations connected by an unspecified LLR. At high T and in the metastable conductive state at low T, X was proposed to be neutral, so as not to compensate the doping and as is consistent with the usual attribution of deep donor properties to V_{As}. Lang et al. (1979) assumed that the D remained a shallow donor while the complex became neutral via the LLR. This implies the X changes from being neutral to negatively charged in n-type material, i.e., that the LLR converts the deep donor to a deep acceptor, and this serves to compensate the sample to a large degree. Van Vechten (1985) argued that this proposal is correct as far as it goes and that the LLR is simply nearest neighbor hopping;

$$D^+ V_{As}^0 + 3e_c^- \langle==\rangle D^+ M_{As}^{2-} V_{Ga}^- \qquad (1)$$

where e_c^- denotes an electron in the conduction band and M may be either Al or Ga, one of the nearest neighbors to the V_{As} that hopped in the reaction.

There have been several other origins proposed for the DX phenomena. (The literature is so extensive that space does not permit one to do justice to all authors. Please consult the review by Lang (1985) and the paper of Gibart et al. (1988).) We divide these into three classes: i) there is no X and there is no LLR but there is a mainly electronic reconfiguration of the isolated donor giving rise to the two states; ii) there is no X but the isolated donor undergoes a LLR itself; iii) there is a LLR and an X but the LLR is not reaction (1) and X is not the V_{As}.

Before distinguishing between (1) and the three classes of alternatives, let us note that all proposed explanations must be similar in one respect that is not always appreciated. In order to account for any self compensation effect that appears as a function of T, alloy composition, and pressure, one must consider the balance of free energies between the conducting and non-conducting states (Van Vechten, 1980). For all proposals, there must be some cost of enthalpy to create the states that trap the carriers and this cost is paid back by the free energy released when carriers fall from the Fermi Sea into these states. That free energy is a function of T, alloy composition, and pressure. Now, the observation of DX behavior in pure GaAs under pressure (Mizuta et al., 1985), when it is not found at 1 atm., has sometimes been claimed to prove there is no X. This would only be valid if one could also show there were not sufficient X in pure GaAs. In fact, positron annihilation studies (Dannefaer et al., 1986) show there are more than sufficient vacancies in all types of GaAs to account for the observed, high pressure DX behavior if (1) is the mechanism. Moreover, Theis et al. (1988) have shown that the DX center <u>is</u> present in unalloyed GaAs, but it there has its ionization level

above the conduction band minimum, so that the level can be occupied and the DX phenomenon can be observed only by hot or highly degenerate electrons. For all proposals, the increase in the GaAs bandgap with pressure increases the payback of trapping the carriers once the deep levels have been created to the point that, at sufficient pressure, the compensation reaction becomes favorable, as it is in the alloy. Moreover, the observed insensitivity of the DX phenomena to the preparation of the sample does not serve to distinguish between alternatives; if, e.g., one varies the As overpressure by a factor of 10 during growth, then one would indeed vary the equilibrium concentration of neutral V_{As}, $[V_{As}^0]$ by a factor of 10, but not $[(V_M M_{As})^{-3}]$, which is a function not only of $[V_{As}^0]$ but also of the Fermi level, E_f. A shift of E_f by only $(kT/3)\ \ln\ 10 = 0.768\ kT$ would compensate this (large) variation of the vapor pressure and likely never be noticed.

We now consider the attributes of reaction (1) and compare and contrast them with the three classes of alternatives:

First, (1) implies that in the non-conducting state most of the donors are undistorted and isolated, i.e., unpaired from X. All D would be compensated if there are 1/3 as many V_{As} performing (1) as there are D. Also, the V_{As} do not need to be near any D to affect the compensation. As assumed implicitly by Lang et al. (1979), there is a tendency for V_{As}^0 to pair with D^+ in well annealed samples, but the attraction is not strong; it derives from the coulomb attraction of the right hand side configuration in (1), which is not the preferred configuration at diffusion-processing T's, for the D^+ and from weaker binding between D and the host atoms which are its nearest neighbors, that allows them to hop away more easily than can host atoms bonded to the other host atoms. Indeed, Mössbauer studies of ^{119}Sn implanted GaAs (Weyer et al., 1980; Bonde-Nielsen et al., 1985) identify a configuration of

the Sn Mössbauer isotope with an isomer shift (IS) of 2.8 mm/s, as contrasted with the 1.8 mm/s IS of the isolated Sn_{Ga} and the 1.9 mm/s IS of isolated Sn_{As}, as the Sn_{Ga}–V_{As} nearest neighbor complex and find these can be dispersed by quenching the sample from T's as low as the 200 - 300 C range. Class 2 alternative theories require every D which participates in the DX phenomena, more than 95% for alloys with 24% to 74% Al, to be in the distorted configuration while class 1 theories require that none of the D be distorted. Class 3 alternatives are not clear as to what they predict but they generally imply a coupling between each D participating and their particular X.

Second, (1) implies the "DX" ought not to be a unique complex but a family of related complexes. This is because vacancy diffusion in III–V crystals occurs largely by nearest neighbor hopping, which requires the formation of pairs of antisite defects and results in a distribution of these normally neutral dipolar defects about the vacancy complexes (Van Vechten, 1975, 1980, 1982, 1984; Van Vechten and Wager, 1985; Juang et al., 1988). Indeed, Bonde–Nielsen et al. (1985) remarked that their studies of the Sn_{Ga} – V_{As} complex found a distribution in electric field gradients that indicates the complex is not unique but subject to disorder. Class 1 and 2 alternatives assert that D is isolated and imply a unique configuration for both states. Class 3 proposals are either based on observations leading to the conclusion that DX is a family of complexes (Farmer et al., 1988) or might imply this for similar reasons. An exception would be the hypothesis that X is a divacancy (Wager, 1988) because divacancies migrate mainly by passing one atom across their two sites, which does not result in antisite defect pair formation.

Third, (1) implies that the fraction (about 1/3 or less) of conductive state donors that are paired with X will be in the field of a 3-fold ionized acceptor in the non-conducting state. Class 1 and 2 alternatives assert the

defect to be neutral, or at most singly ionized, in the non-conducting state. The Class 3 alternative of Wager (1988) asserts DX to be neutral; Farmer et al. (1988) do not answer the question.

Fourth, for the case of Sn_{III} doping, which was used for the recent Mössbauer experiments (Gibart et al., 1988) we discuss, it follows that (1) implies the corresponding IS will be relatively large. Accurate and reliable ab initio IS calculations are available only for perfect substitutional sites (Svane, 1988) but trends can be estimated with the aid of empirical identifications (Shenoy and Wagner, 1978; Bonde-Nielsen, 1985; Williamson, 1986, 1987). It is well established that the IS is directly proportional to the electron density at the nucleus of the Mössbauer isotope, so, e.g., the difference of the IS's between Sn_{Ga} and Sn_{As}, 1.8 vs. 1.9 mm/s, corresponds to the charge transfer between these two sites imposed by the partly ionic character of GaAs. The IS = 2.8 mm/s identified with Sn_{Ga}-V^0_{As} in GaAs is attributed to the redistribution of charge from the bonds dangling into the vacancy cavity, mainly to the nearest neighbors and largely to states that overlap their core much more than do covalent bonds. It seems that those engaged in the identification of Mössbauer lines with defects have previously not considered antisite defects the among possible candidates, despite the firm establishment of their prevalence in III-V materials and participation in other metastable deep level defects such as "EL2" in GaAs and "M" in InP. (Cf., e.g., Van Vechten, 1980; Wager and Van Vechten, 1987; Juang et al., 1988 and many references therein.) While a shallow impurity (dopant) in the vicinity of the Mössbauer isotope would have small effect on IS due to the broad hydrogenic orbit of the carrier they bind, deep levels such as the Ga^{2-}_{As} must make a significant contribution by binding two electrons in a highly localized state that would be a nearest neighbor to a Sn_{Ga} donor if (1) occurs

from a V_{As} paired with it. The V_{Ga}^{-} deep level on the second neighbor site of the Sn_{Ga} should also contribute. Thus, we expect the IS for $Sn_{Ga}^{+} - M_{As}^{2-} - V_{Ga}^{-}$ to be even larger than identified for $Sn_{Ga}^{+} - V_{As}^{0}$, about 3.1 mm/s judging by the trend cited. We note that the only previously identified source of an IS > 3.0 mm/s for ^{119}Sn in GaAs is the Sn interstitial produced by the decay of ^{119}Cd which gives IS = 3.4 mm/s (Bonde-Nielsen et al. 1985). Interstitial Sn is highly unstable due to its very high enthalpy of formation, which results from the same large overlap of host bond charge density with its core (Van Vechten 1980), and almost certainly resulted from the 20 eV recoil energy of the Cd decay.

Reaction (1) also implies that the fraction of Sn donors which are paired with the vacancy-antisite complex are in the presence of a strong electric field gradient which should produce a significant quadrupole splitting of the Mössbauer spectrum. The author is not aware of an accurate method to estimate the magnitude of this quadrupole splitting of the IS and thus to predict whether or not it should be resolved experimentally.

Fifth, (1) implies the net charge of the DX complex in the non-conductive state to be -2. As the D^{+} is either first or second neighbor to the V_{Ga}^{-} for the majority of cases in well annealed samples, spin pairing is expected. Thus, for the type of samples generally investigated, the DX complex should have no unpaired spin and should be paramagnetically inactive. Class 1, 2 and 3 alternatives generally imply DX should be EPR active in the non-conductive state; there are a few exceptions.

We now consider the Mössbauer data recently reported by Gibart et al. (1988), who concluded the spectra imply LLR (contradict class 1 alternatives), but did not express a view what they imply re antisite defects, vacancies, etc. (Mooney et al. (1988) also affirm the presence of LLR.)

The samples were an alloy series of metal-organic vapor phase epitaxially (MOVPE) grown layers ranging from pure GaAs to pure AlAs doped during growth with ^{119}Sn. Thus, there was no ion implantation and no transmutation of the Mössbauer isotope. All measurements were at T = 76 K in the dark so that samples with Al mole fraction, x, between 0.25 and 0.75 must have been in the non-conductive state. Three Mössbauer lines were noted and traced through the alloy series. Table 1 reproduces relevant data. The variation of IS with x is moderately greater than experimental uncertainty, which is greatest (0.2 mm/s) for line 2 at x = 0.2 and typically 0.1 mm/s or better. The intensity, F, of line 1 (Sn_{III}) and of line 3 vary with x much more than does that of line 2. For GaAs Williamson (1986, 1987) identified line 2 with neutral Sn agglomeration (likely an artifact of the growth process, e.g., balling of Sn on the sample surface). He also proposed that line 3 in pure GaAs is a quadrupole doublet of line 2. However, one might not expect to resolve the doublet if disorder broadening is as severe as Bonde-Nielsen et al. (1985) found. In any case a quadrupole doublet would have to maintain a fixed ratio, which is generally expected to be 1.0, of intensity from one sample to the next. (Agglomerations of the dopant element cannot participate in the DX phenomena but they cannot be eliminated from the observed spectrum either.)

The third line appeared weakly with an IS about as large as that of the Sn interstitial in the x = 0 and x = 1 samples, where DX behavior is not evident unless hot electrons are injected, or the doping is highly degenerate, or pressure of order 20 kbar is applied. Note that Sn interstitials are not expected as there was no transmutation recoil nor any ion implantation to produce them. In the range of x where DX behavior is dominant, roughly $0.24 < x < 0.74$, line 3 is second in strength and displayed an IS of about 3.1 mm/s. Gibart et al. (1988) suggest that DX appears as a doublet, part of both

lines 2 and 3, the remainder being the Sn agglomerations. We question this suggestion because the F ratio does not appear to be constant and because there is evidence for disorder about DX from Farmer et al. (1988). Alternatively, we suggest: a) doublets are not resolved here; b) line 2 alone results from agglomerations only; and c) line 3 results from complexes based on Sn_{III}-V_{As} <==> Sn_{III}-III_{As}-V_{III}, which are expected to occur also for alloy compositions where the DX transition is not favored (Van Vechten 1985).

TABLE 1

Values of the fractional % resonance intensity, F, and of isomer shift, IS, for samples of various mole fraction of Al content, x. Uncertainty of F indicated in parentheses. Quoted from Gibart et al. (1988).

line\x		0.0	0.20	0.30	0.40	0.43	0.70	0.76	1.00
1	F	73(3)	75(4)	40(7)	42(5)	50(4)	65(6)	67(5)	69(5)
	IS	1.78	1.78	1.81	1.89	1.74	1.76	1.81	1.75
2	F	17(3)	14(3)	28(7)	24(11)	28(4)	14(8)	21(6)	17(6)
	IS	2.5	2.22	2.39	2.57	2.26	2.38	2.47	2.53
3	F	10(3)	11(3)	32(7)	34(11)	22(3)	21(5)	12(3)	14(3)
	IS	3.30	3.26	2.94	3.05	3.02	3.02	3.57	3.48

It is plain that, if we take the F of line 3 alone to be DX, these data agree with (1) on the first point; in the range of the DX phenomena there are roughly twice as many isolated Sn_M normal donors as "DX complexes" in the non-conducting state. The conclusion that the non-conducting state still has a majority of donors in their normal (shallow) configuration, so that the DX complexes must be multiple acceptors, is also in accord with recent FTIR studies (Murray et al., 1988) and with EXAF studies (Mizuta and Kitano, 1988).

If, on the other hand, we assume that Gibart et al. are correct that DX appears as a doublet, part of both lines 2 and 3 which also contain contributions from Sn agglomeration, then we should analyze the data as in Table 2. There we show the intensities of the three lines relative to their values in unalloyed GaAs, $F_i(x) - F_i(0)$, for the various values of x, the mole fraction of AlAs in the alloy, and then calculate the ratio of DX centers to total donors as

$$[DX]/[D] = \frac{(F_2(x) - F_2(0)) + (F_3 - F_3(0))}{F_1(0)} \quad . \tag{2}$$

For completeness and comparison, we also show the ratio of the concentration of DX to total donors calculated according to our suggestion that DX contributes only to line 3, where its quadrupole splitting is unresolved;

$$[DX]/[D]^* = \frac{F_3(x) - F_3(0)}{F_1(0)} \quad . \tag{3}$$

It seems safe to assume that the true situation lies between these two propositions.

For the four samples wherein DX is expected to appear (x = 0.30, 0.40, 0.43 and 0.70), Eq. (2) yields an average empirical value of 32%, compared with the value of $\lesssim$33% implied by (1), while Eq. (3) yields an average empirical value of 24%. It must be admitted that for the x = 0.30 and 0.40 samples, Eq. (2) could also be used to support the model of Chadi and Chang (1988), that the DX phenomena is a negative-U behavior of the isolated donors themselves;

$$D^+ + 2e^- \Leftrightarrow D^- \quad ; \tag{4}$$

Table 2

Analysis of the Mössbauer data of Gibart et al. (1988) according to their assumption that DX appears in both line 2 and line 3, and according to our proposal that it contributes only to line 3.

x	0	0.20	0.30	0.40	0.43	0.70	0.76	1.00
$F_1(x) - F_1(0)$	0	2	−33	−31	−23	−8	−6	−4
$F_2(x) - F_2(0)$	0	−3	11	7	11	−3	4	0
$F_3(x) - F_3(0)$	0	1	22	24	12	11	2	4
[DX]/[D] (%)	0	−2.7	45	42	31	11	8.2	5.5
[DX]/[D]* (%)	0	1.3	30.1	32.9	16.4	15.1	2.7	5.5

which implies the ratio should be 50%. Among other reasons, we find the Chadi-Chang model difficult to accept because the DX phenomena occurs for all donors, whether they reside on the anion or in the cation lattice site, at the same concentrations, Fermi level position, temperature etc. In any case it is plain that Mössbauer data are consistent with (1) regarding the third point.

As regards the second point, these Mössbauer data are also in accord with (1) in that they indicate disorder broadening of the DX complex line, as expected for a vacancy related complex and as previously found by Bonde-Nielsen et al. (1985) for the $Sn_{Ga} - V_{As}$ complex in GaAs. There are several other experiments that indicate DX is not a single defect but a family of related defects. These include the sample dependence of the DLTS spectra of DX transitions reported by Farmer et al. (1988) and the studies of DX transition parameters in unalloyed GaAs under pressure by Mizuta et al. (1985). These observations argue strongly against Class 1 and Class 2 alternatives. It must also be remarked phonon scattering studies by

Narayanamurti et al. (1979) gave early and persuasive evidence for the conclusion $X = V_{As}$ in the conductive state.

Finally, regarding the fifth point, one can only say that despite a great deal of searching, no paramagnetic active center has been found that could be ascribed to the DX complex (Weber 1988).This is in accord with (1) and contradicts many alternatives.

We acknowledge the partial support by the U. S. Air Force Office of Scientific Research (AFOSR 86-0309). Discussions of Mössbauer effect and data with D. L. Williamson are gratefully acknowledged.

References

Bonde-Nielsen K, Grunn H, Haas H, Pedersen F T, and Weyer G 1985 13th International Conference on Defects in Semiconductors ed. L C Kimerling and J M Parsey (New York AIME) *J. Electron. Mat. 14a* 1065.

Chadi D J and Chang K J 1985 *Phys. Rev. Lett. 61* 873.

Dannefaer S and Kerr D 1986 *J. Appl. Phys. 60* 591.

Farmer J W, Hjalmarson H P, and Samara G 1988 *Mat. Res. Soc. Symp. 104* 585.

Gibart P, Williamson D L, El Jani B, and Basmaji P 1988 *Phys. Rev. B 38* 1885.

Juang M T, Wager J F, and Van Vechten J A 1988 *J. Electrochem. Soc. 135* 2023.

Lang D V and Logan R A 1977 *Phys. Rev. Lett. 39* 635.

Lang D V, Logan R A, and Jaros M 1979 *Phys. Rev B 19* 1015.

Lang D V 1985 *Deep Centers in Semiconductors* ed. by S T Pantelides (New York, Gordon Beach) p. 489.

Mizuta M and Kitano T 1988 *Appl. Phys. Lett. 52* 126.

Mizuta M, Tachikawa M, Kukimoto H, and Minomura S 1985 *Jpn. J. Appl. Phys. 24* L143.

Mooney P M, Northrop G A, Morgan T N, and Grimmeiss H G 1988 *Phys. Rev B 37* 8298.

Murray R, Newman R C, and Nandhra P S 1988 *Mat. Res. Soc. Symp. Proc. 104* 543.

Narayanamurti V, Logan R A, and Chin M A 1979 *Phys. Rev. Lett. 43* 1536.

Shenoy G K and Wagner F E 1978 *Mössbauer Isomer Shifts* (Amsterdam, North-Holland).

Svane A 1988 *Phys. Rev. Lett 60* 2693.

Theis T N, Mooney P M, and Wright S L 1988 *Phys. Rev. Lett. 60* 361.

Van Vechten J A 1975 *J. Electrochem. Soc. 122* 423.

Van Vechten J A 1980 *Handbook on Semiconductors* Vol. 3 ed. S. P. Keller (Amsterdam, North Holland) p. 1.

Van Vechten J A 1982 *J. Appl. Phys. 53* 7082.

Van Vechten J A 1984 *J. Phys. C: Solid State Phys. 17* L933.

Van Vechten J A 1985 *Mat. Res. Soc. Symp. Proc. 46* 83.

Van Vechten J A and Wager J F 1985 *Phys. Rev. B 32* 5259.

Wager J F and Van Vechten J A 1987 *Phys. Rev. B 35* 2330.

Wager J F 1988 unpublished.

Weber E R 1988 unpublished.

Weyer G., Damgaard S, Petersen J W, and Heinemeier J 1980 *J. Phys. C:Solid State Phys. 13* L181.

Williamson D L 1986 *J. Appl. Phys. 60* 3466.

Williamson D L, Gibart P, El Jani B, N'Guessan K 1987 *J. Appl. Phys. 62* 1739.

TEMPERATURE DEPENDENCE OF BAND OFFSETS FOR HETEROJUNCTIONS IN GENERAL AND FOR THE PARTICULAR CASES OF AlAs-GaAs AND HgTe-CdTe

J.A. Van Vechten

Center for Advanced Materials

Department of Electrical and Computer Engineering

Oregon State University, Corvallis, OR 97331

and

K.J. Malloy

Air Force Office of Scientific Research

Directorate of Electronic and Material Sciences

Bolling Air Force Base, DC 20332

ABSTRACT

We devise a simple theory for the temperature dependence of the valence and conduction band offsets in semiconductor heterojunctions using the thermodynamic point of view. The temperature dependencies of the offsets originate from the separate contributions of holes in the valence bands and of electrons in the conduction bands of the two semiconductors to the temperature dependence of their respective band gaps. We use the determination of these contributions previously made by Heine and Henry from isotope shifts of luminescent lines due to impurities. By considering this temperature dependence of the band offset we suggest an explanation for the discrepancy between the determinations of AlAs-GaAs valence band offset by Wolford et al. and by Batey and Wright. Whereas for most pairs of semiconductors the bands move in the same direction with varying temperature, for the particular case of

HgTe–CdTe, they move in opposite directions. From this we predict a much greater than usual temperature dependence for the band offsets for HgTe–CdTe junctions and reconcile the major discrepancy between valence band offsets determined by Kowalczyk et al. and by Chow et al.

INTRODUCTION

With the dramatic advance of semiconductor heterojunction technology[1-10] has come a concomitant increase of interest in the origin[11-18] of the valence and conduction band offsets, or discontinuities, ΔE_v and ΔE_c, respectively. In spite of the implied simplicity in the concept of band offsets, no single theoretical description has been acknowledged as correct. To add an additional factor to this uncertainty, we wish to explore the temperature dependence of the band offsets, an issue ignored until raised by recent experiments.[19,20]

Approaches to the problem of the band offsets divide into two categories: thermodynamic and electronic structure calculations. In addition to differences in the method of calculation between the two approaches, there is a difference in the way they view the small, but detectable, effect of depolarizing fields set up around the boundaries of the samples as a result of dipole layers that are concomitant with those boundaries. The distinction is particularly clear and acute for the case of the band offset to vacuum, i.e., for the questions of the "work functions" or "ionization potentials" of metals or of semiconductors. For the purposes of the present paper, which is concerned with the TEMPERATURE DEPENDENCE of the band offsets, the point to keep in mind is that the effect of depolarizing fields on junctions between two semiconductors is small at all temperatures and, owing to its origin, should be expected to have very little temperature dependence in the range $0 < T < 310$ K in which the data we discuss was taken. However, there seems to be such a semantic disagreement, and some confusion, between the two camps over this issue that some space must be taken to discuss it. (References 11 and 13 contain previous attempts to resolve these semantic arguments and confusion.)

Those who use thermodynamic approaches are usually concerned with differences in free energy, chemical potential, etc., between bulk phases. For such a consideration, any junction between two phases is irrelevant; there need not be any junction so long as they can somehow exchange the energy, particles, etc. relevant to the problem. Thus, with the thermodynamic approach to the band offset (and work function) problems[11-13,15], one ascribes absolute values (usually relative to an idealized "vacuum level"[11]) to the enthalpies, entropies, and standard free energies of the valence and conduction band edge density of states distributions in the equipotential bulk region of each semiconductor,[21,22] when no bias field is applied externally. (Recall that the standard free energies are the total free energies minus all explicitly concentration dependent entropy terms; in particular, the entropy terms resulting from the statistical distribution of the free carriers among the band states are absent from these standard free energies. Thurmond has given a particularly clear and authoritative account[21] of this point. Any good text book contains a general discussion, usually when treating the Law of Mass Action.) Those who use this thermodynamic approach then DEFINE the banc offset as simply the difference between these absolute standard free energies.

$$\Delta E_V(A/B) = E_V(A) - E_V(B) \tag{1}$$

$$\Delta E_C(A/B) = E_C(A) - E_C(B) \tag{2}$$

The result is, by this definition, independent of crystallographic orientation and transitive from one material to another; i.e., e.g.,

$$\Delta E_C(\text{AlAs/GaAs (110)}) = \Delta E_C(\text{AlAs/GaAs (100)}) \tag{3}$$

and

$$\Delta E_v(AlAs/GaP) = \Delta E_v(AlAs/GaAs) + \Delta E_v(GaAs/GaP) \ . \quad (4)$$

Furthermore, for any pair of semiconductors A and B,

$$\Delta E_v(A/B) - \Delta E_c(A/B) = \Delta E_{cv}(B) - \Delta E_{cv}(A) \ , \quad (5)$$

where

$$\Delta E_{cv} = E_c - E_v \quad (6)$$

denotes the band gap, E_v is the absolute standard free energy of the valence band edge distribution, and E_c is the absolute standard free energy of the conduction band edge distribution.

The fact that, by the thermodynamic approach definition, band offsets must be independent of orientation and transitive can readily be appreciated by considering, as in Fig. 1, the addition of a third material, C, to form junctions to the first two, A and B. The change in energy, enthalpy, total free energy or standard free energy upon transferring one carrier from the bulk of A to the bulk of B must be independent of the path utilized. Thus, the transit may be directly through the A/B junction or through the A/C junction followed by a path through sample C and through the C/B junction. Because this is true regardless of the orientation of either junction and of the material C, which might be vacuum, it is plain that these bulk-to-bulk band offsets must be transitive and independent of orientation.

Now, it is well known[11] that when one measures, e.g., by photoemission, the quantity generally known as the "work function" of a metallic or semiconducting crystal, the result is depends upon crystal orientation to a small degree; it depends very sensitively on the presence any foreign atoms, e.g., Cs or O, on the surface. It is also well known[11] that the reason for this is

that such experiments do not measure the chemical potential difference, etc, between the bulk of the crystal and the vacuum level, but instead they measure an effective potential energy barrier between a region in the sample near the surface and a region close above the sample and because the termination of the sample at surfaces of various orientation, chemistry and condition produces electrostatic dipoles that depend upon that orientation, chemistry and condition. These dipoles produce depolarizing fields that extend around the exteria of the sample, and, to a much smaller degree, extend also within the sample near the surfaces (or junctions). These fields would be irrelevant to the experiment if the electron were making a transition from the bulk to the idealized point at infinite separation from the sample and from all other charges, but practical experiments do not work that way.

Those who use the thermodynamic approach realize, and readily admit, that practical experiments which measure the "work function" do find a quantity which is neither transitive nor orientation independent. They distinguish[11-13] the "work function" from the thermodynamic parameter that they would call the "band offset between the sample material and vacuum." Some attempt[13] to avoid confusion on this point by calling the thermodynamic "band offset to vacuum" the "ionization potential." It is quit a tractable problem to extract[11] the "ionization potential" from a adequate set of data on the "work function."

Unless the sample is heated so much that its surface chemistry changes or the surface atoms diffuse enough to change their local surface array, the dipole layer fields that produce this distinction between "work function" and "ionization potential" will not change. If the surface dipoles do not change with T, then the magnitude of the distinction will not change with T. Thus,

the dipole fields will then make no contribution to any measured T dependence of either the "work function" or the "ionization potential."

In general, practical junctions between crystalline samples also produce electrostatic dipoles that also produce fringing, or depolarization fields. For semiconducting samples these fields are much weaker than those in vacuum because of the conductivity of the two solids. However, they do have finite effect upon the effective barrier a carrier encounters transiting a the junction in a practical experiment. Just as the chemistry of the surface has much more effect on the "work function" than does the crystal orientation, the effective barrier at a semiconductor junction is expected to be much more sensitive to the local chemistry of the interface than to its orientation. (Indeed, the Freeouf-Woodall and Spicer's defect model for Schottky barrier heights attribute dramatic effect to a local chemistry at the interface different from that of the bulk sample.) Again, those with a thermodynamic point of view regard these junction field effects as a minor nuisance that should be extracted from the raw data in order to determine the true, bulk-to-bulk band offset, which must be transitive and orientation independent.

With the electronic structure approach,[14,16-18] one solves a Schrödinger equation for the particular chemistry and orientation of the junction as exactly and as rigorously as possible. With such an approach one has to deal explicitly with the junction dipole in detail. Those who use this approach regard these dipoles and their depolarization fields with more respect than do those who use the thermodynamic approach. They generally include their effect when they seek to relate the results of their Schrödinger equation calculation, which are eigenvalues and not free energies, on the two sides to ΔE_v and ΔE_c. (The problem of relating eigenvalues to the enthalpies and free energies

of the band density of states distributions is discussed on page 8 to 18 of Ref. 22.)

Thus, with the electronic structure approach, one includes the small effect of the depolarizing field in ones definition of the "band offset" which is then an effective barrier height for transit of the carrier. Fortunately, the two approaches are now beginning to concur[18] with one another in cases like AlAs/GaAs where the close matching of the lattice constants make for very simple junctions.

The effective barrier definition of the "band offset" depends on the sample geometry and the distance of the two reference points from the interface. In general, it can be non-transitive and orientation dependent. It may be claimed to be more directly related to the result of practical experiments, but, in our view, it is not nearly so well nor clearly defined a concept and it is not nearly so easily applied to the question of the temperature dependence of the band offsets. Again, the reader should keep in mind that, because these junction dipole fields should not depend on T in the range of the relevant experiments, this distinction between the definitions of "band offset" will have no significance for any discussion of the <u>temperature dependence</u> of the band offset.

The purpose of the present work is to address the problem of the temperature dependence of the band offsets. Although it seems not to have been done before, we will see that this is a very simple problem if one uses the thermodynamic approach. (It would appear to be rather difficult from the electronic structure approach.) We see from Eqs. (1) to (6) that all that is required is a determination of the variation with T of E_v and E_c for the two semiconductors that form the junction. Most of what is required to establish

$E_v(T)$ and $E_c(T)$ has already been developed and published with reference to the closely related problem of $\Delta E_{cv}(T)$, for which empirical values are established for most semiconductors.[21]

Also for the problem of $\frac{d\Delta E_{cv}}{dT}$ there is a simple thermodynamic approach,[22,23] which avoids the difficulties of electronic structure calculations. One notes that the explicit effect of T on the electronic eigenvalues is a very small part of the total $\frac{d\Delta E_{cv}}{dT}$ (as can be determined from the coefficient of thermal expansion and the pressure dependence of ΔE_{cv}) and the vast majority of the effect comes from the electron-phonon interaction. The electron-phonon interaction can be evaluated by considering the small perturbative effect upon the phonons of the thermally excited electrons, e_c's, and holes, h_v's, which are never more[21] than 1 part in 10^4 of the bonding electron density, using a simple bond-charge theory[24,25] of phonon energies. This approach has been shown to give quantitative agreement with experiment for the phonon frequencies,[26] as well as for $\Delta E_{cv}(T)$, up to the melting points of Si and Ge. (Of course, there is also an electronic structure approach,[27] which considers the massive perturbation of (vibronic) electronic levels by phonons, which are several times more numerous than are host atoms. These two approaches now concur rather well.)

TEMPERATURE DEPENDENCE OF BAND OFF-SET

With the thermodynamic approach[21,22] one readily sees that for most semiconductors ΔE_{cv} decreases with increasing T because the thermal excitation of both e_c's and h_v's soften transverse acoustic phonon modes by (respectively) putting charge into antibonding states in the conduction band and taking it away from bonding states in the valence band. This results in a large

positive (standard) entropy, ΔS_{cv}, for the reaction[21,22] that thermally excites e_c's and h_v's;

$$0 \rightarrow e_c + h_v \ , \ \Delta E_{cv} \ . \tag{7}$$

As ΔE_{cv} is the (standard) chemical potential for $e_c + h_v$ pairs, it is equal to a free energy and follows the universal relation[21] of Gibbs

$$\Delta E_{cv}(T) = \Delta H_{cv}(T) - T\Delta S_{cv}(T) \ , \tag{8}$$

where ΔH_{cv} is the corresponding (standard) enthalpy of the reaction. $\Delta E_{cv}(T)$, $\Delta H_{cv}(T)$, and $\Delta S_{cv}(T)$, as determined by Thurmond in Ref. 21, for the case of GaAs are plotted in Fig. 1. Note that basic considerations of thermodynamics require that $\Delta S_{cv}(T=0K) = 0$, and if $\Delta S_{cv} > 0$ for $T > 0$, then $d\Delta E_{cv}(T)/dT < 0$ and $d\Delta H_{cv}(T)/dT > 0$.

Whereas ΔS_{cv} denotes the combined effect of one e_c and one h_v, it is sometimes possible to separate the two effects by observing the isotope shifts of zero-phonon optical transitions due to impurities.[28,29] One may deduce from the isotope shift the local softening effect on the phonon modes about the impurity when one electron or one hole localizes into a state about the impurity. Heine and Henry studied this problem[28] for several semiconductors, particularly GaP and ZnO. They concluded that, at least for these cases, a hole is almost four times as effective in softening phonons as is an electron. That is

$$S_h = (3.5 \pm 1) \ S_e \ , \tag{9}$$

with

$$\Delta S_{cv} = S_e + S_h \ . \tag{10}$$

As suggested by Heine and Henry, we assume that this distribution of the weight of the two contributions is essentially the same for all tetrahedrally bonded semiconductors.

Let us now define the band offset problem in thermodynamic terms. (Care is required not to make a sign error.) We have the two semiconductors, A and B, and consider the reaction that transfers an e_c from B to A at the cost of free energy $\Delta E_c(A/B)$,

$$A{:}0\ ,\ B{:}e_c^-\ \Rightarrow\ A{:}e_c^-\ ,\ B{:}0\ \ ,\ \ \Delta E_c(A/B) = E_c(A) - E_c(B)\ . \tag{11}$$

We also have the reaction that transfers a h_v from B to A at the cost of free energy $-\Delta E_v(A/B)$ (note the effect of the difference in the sign of the charge of the h_v and the e_c),

$$A{:}0\ ,\ B{:}h_v^+\ \Rightarrow\ A{:}h_v^+\ ,\ B{:}0\ \ ,\ \ -(\Delta E_v(A/B) = E_v(A) - E_v(B))\ . \tag{12}$$

The vacuum level, or any other reference state, R, may be introduced by breaking these reactions into two parts, i.e.,

$$B{:}e_c^-\ ,\ R{:}0\ \Rightarrow\ B{:}0\ ,\ R{:}e^-\ \tag{13}$$

followed by

$$R{:}e^-\ ,\ A{:}0\ \Rightarrow\ R{:}0\ ,\ A{:}e_c^-\ , \tag{14}$$

and correspondingly for the hole reaction. Adding reactions (11) and (12), we have

$$A{:}0\ ,\ B{:}e_c^- + h_v^+\ \Rightarrow\ A{:}e_c^- + h_v^+\ ,\ B{:}0\ \ ,\ \ \Delta E_c(A/B) - \Delta E_v(A/B)\ , \tag{15}$$

which immediately implies

$$\Delta E_{cv}(A) - \Delta E_{cv}(B) = \Delta E_c(A/B) - \Delta E_v(A/B) \ , \tag{16}$$

as claimed above at Eq. (5). When we differentiate Eq. (16) with respect to T, we have

$$\frac{d\Delta E_{cv}(A)}{dT} - \frac{d\Delta E_{cv}(B)}{dT} = \frac{d\Delta E_c(A/B)}{dT} - \frac{d\Delta E_v(A/B)}{dT} \ . \tag{17}$$

We now recall the basic thermodynamic identity that the entropy is the negative of the derivative of the corresponding free energy (or chemical potential) with respect to T, so that

$$S_c(A) = - \frac{dE_c(A)}{dT} \quad ; \quad S_v(A) = - \frac{dE_v(A)}{dT} \ , \tag{18}$$

and, of course, the same for B while for the band gaps themselves

$$\Delta S_{cv}(A) = - \frac{dE_{cv}(A)}{dT} \ . \tag{19}$$

Finally, when we compare Eq. (10) with (18), we have the interesting result that

$$S_e + S_h = S_c - S_v \ , \tag{20}$$

or

$$S_c = \Delta S_{cv} + S_h \ . \tag{21}$$

This merely states that fact that, in order for the band gap to decrease with rising T, the absolute free energy E_c must fall faster than the absolute free energy E_v by ΔS_{cv}. This is illustrated in Fig. 2 for the cases of GaAs and AlAs.

We now introduce the physical assumption of Eq. (9), that the fraction of ΔS_{cv} due to the h_v is a constant 77 (± 5) % for all tetrahedrally bonded semiconductors, and we equate this with S_v.

$$S_v = S_h \ , \tag{22}$$

so that

$$\frac{d\Delta E_v(A/B)}{dT} = 0.77(\pm\ 0.05)\left[\frac{d\Delta E_{cv}(A)}{dT} - \frac{d\Delta E_{cv}(B)}{dT}\right] . \tag{23}$$

While Eq. (22) is attractive on intuitive physical grounds, it is not clear that it is rigorously exact. The biggest problem is that it assumes that the effect on the lattice modes of a localized hole or electron, as observed in the isotope shift experiments,[28] is the same as that of a delocalized h_v or e_c in the band edge density of state distribution. Van Vechten and Thurmond have argued that this should be so to a good approximation[22,29] and correlated it to the proposition that for any tetrahedral semiconductor the temperature dependence of the various direct band gaps in the optical spectrum (the fundamental gap and all higher gaps) should be the same. (This implies that a delocalized state at any point in the Brillouin zone has the same effect on the lattice so that the effect of a localized state, having components from many points in the zone, would also be the same.) What experiments have been done to test this hypothesis[30] (on Si), support it. Aside from this problem of a possible distinction between localized and delocalized carriers, one can question the consequences of the possible alternative assumptions that $S_v < S_h$ or $S_v > S_h$. If $S_v < S_h$, one would have to conclude that the addition of antibonding charge, i.e., the e_c, somehow stiffens the lattice relative to the degree indicated by the isotope shift experiments, so that the entropy of the

valence band edge distribution could be reduced. This seems most unlikely. If $S_v > S_h$, then there must be some contribution to the entropy of both the valence and conduction band density of states beyond that due to the lattice modes. (Recall the quantitative description of the variation of the mode frequencies[26] to the melting points of Si and Ge.) No source of such a contribution is apparent.

Let us now note the consequence of Eqs. (21), (22) and (23) for ΔE_c;

$$\frac{d\Delta E_c(A/B)}{d\bar{T}} = 1.77(\pm 0.05)\left[\frac{d\Delta E_{cv}(A)}{d\bar{T}} - \frac{d\Delta E_{cv}(B)}{d\bar{T}}\right] . \tag{24}$$

Furthermore, because ΔH_v and ΔH_c are connected by thermodynamic identities[21] to the ΔS's and ΔE_c and ΔE_v;

$$\frac{d\Delta H_v(A/B)}{d\bar{T}} = 0.77(\pm 0.05)\left[\frac{d\Delta H_{cv}(A)}{d\bar{T}} - \frac{d\Delta H_{cv}(B)}{d\bar{T}}\right] \tag{25}$$

and

$$\frac{d\Delta H_c(A/B)}{d\bar{T}} = 1.77(\pm 0.05)\left[\frac{d\Delta H_{cv}(A)}{d\bar{T}} - \frac{d\Delta H_{cv}(B)}{d\bar{T}}\right] \tag{26}$$

To carry this discussion any further, we must consider specific cases. We find data adequate for a discussion available for two cases, those of GaAs–AlAs junctions and of HgTe–CdTe junctions, and that these generally support the foregoing simple theory.

THE CASE OF AlAs – GaAs

In the AlAs–GaAs heterojunction, E_v for GaAs is higher[31] than for AlAs at T=0. It was found in Ref. 32 that $\frac{d\Delta E_{cv}}{d\bar{T}}$ is a linear function of the mole fraction of Al in the $Al_xGa_{1-x}As$ alloy system for $x < 0.50$. We use their

extrapolation to AlAs, which concluded that $\frac{d\Delta E_{cv}}{dT}$ for AlAs is 1.6 ± 0.2 times that for GaAs. Consequently, E_v(GaAs) will decrease with rising T slower than will E_v(AlAs) and ΔE_v(GaAs/AlAs) will become larger. (See Fig. 2.) However, ΔH_v(GaAs/AlAs), which equals ΔE_v(GaAs/AlAs) at T = 0K, will decrease with rising T because H_v(GaAs) will increase slower than H_v(AlAs).

Consider now what empirical information relevant to the simple theory of Eq. (23) and (24) is available. The theory implies that for most cases $\frac{dE_v(A/B)}{dT}$ and $\frac{dE_c(A/B)}{dT}$ are both small because the $\frac{d\Delta E_{cv}(A)}{dT} \sim \frac{d\Delta E_{cv}(B)}{dT}$ for most pairs of semiconductors A and B. Furthermore, the total variation of the ΔE_{cv}(A) over the range of accurate experiments, generally from T = 0K to 300K, is not much larger than the experimental uncertainty in ΔE_v(A/B) or ΔE_c(A/B). (For GaAs ΔE_{cv}(T=0 K) − ΔE_{cv}(T=295 K) = 94 meV while ΔH_{cv}(T=295 K) − ΔH_{cv}(T=0K) = 39 meV.) This is certainly consistent with the fact that the authors have not been able to find any explicit discussion of the temperature dependence of band offsets in the previous literature. However, one can look at discrepancies in determinations made at different T's.

The case of AlAs–GaAs heterojunctions has been studied probably more carefully and with better prepared samples than any other. For the cases $Al_xGa_{1-x}As$ alloys on GaAs, Wolford et al. determined[31] that ΔE_v(T=8K) = 110 ± 8 meV for GaAs–$Al_{0.28}Ga_{0.72}As$ and ΔE_v(T=8K) = 320 ± 10 meV for GaAs–$Al_{0.70}Ga_{0.30}As$. They used optical methods; it is shown in Ref. 22 that such optical experiments do in fact measure chemical potentials and thus, as Wolford et al. imply, do determine ΔE_v's or ΔE_{cv}'s rather than ΔH_v's or ΔH_{cv}'s. Another careful determination on similarly well prepared samples was made[33] at higher T's by Batey and Wright. They studied the thermionic emission of h_v's across

the junction in p-type material as a function of alloy composition and of T for the interval 79.6K to 294.4K. They concluded

$$\Delta E_v(GaAs/Al_xGa_{1-x}As) = 0.55x \text{ eV } (\pm 20 - 40 \text{ meV}) \quad (27)$$

where x is the mole fraction of Al in the alloy. This implies ΔE_v = 154 ± 30 meV for GaAs–$Al_{0.28}Ga_{0.72}As$ in contrast to Wolford et al.'s 110 ± 8 meV and 385 ± 30 meV for GaAs–$Al_{0.70}Ga_{0.30}As$, in contrast to 320 ± 10 meV.

Batey and Wright obtained Eq. (27) by fitting their data for the thermionic emission current, J_s, with a Richardson equation,

$$J_s = A^*T^2\exp(-\phi/kT) \ , \quad (28)$$

where A* is the Richardson constant and s the activation barrier for holes. The limit of ϕ as bias voltage goes to zero can be simply related to ΔE_v, if, as Batey and Wright did, account is taken of the temperature variation of Fermi level. However, in their analysis of Arrhenius plots of versus 1/T, ΔE_v was treated as if it were independent of T. A reanalysis of the raw data including the expected variation of ΔE_v, and thus of ϕ, with T is required.

Although the issue may now be unresolvable, we can discuss the implications of the temperature dependence of ΔE_v regarding a reconciliation of these two experiments. First let us note that, since ϕ is a free energy barrier, as is ΔE_v, Batey and Wright should have measured a larger value than Wolford et al. simply because, as noted above, the facts that $\frac{d\Delta E_{CV}}{dT}$ is greater for AlAs than for GaAs and that $E_v(GaAs) > E_v(AlAs)$ imply $\Delta E_v(GaAs/AlAs)$ increases with T. For the case of the 70% AlAs alloy, we have

$$\Delta E_{CV}(T=0K) - \Delta E_{CV}(T=295K) = 94\ [1 + 0.7\ (0.6 \pm 0.2)] \text{ meV}$$

$$= 134 \pm 13 \text{ meV} \ . \quad (29)$$

Thus, ΔE_{cv} decreases 40 ± 13 meV more in the alloy than in pure GaAs and, by our argument, 77 ± 5% of this difference occurs at the valence band edge. Thus, we predict ΔE_v should be as much as 31 ± 10 meV greater at 295K than at OK. Depending on the details of Batey and Wright's analysis of their data, then one sees that the expected variation of ΔE_v with T accounts for 31 ± 10 meV of the 65 ± 40 meV discrepancy between that determination and the optical experiment[31] at a constant T = 8K.

While a final resolution of this discrepancy will require a reanalysis of the raw data of the thermionic emission study, which is not available to the present authors. However, it would appear that the consequences of the variation of ΔE_v with T have the approximate magnitude and arguably the correct sign to reconcile these two fine experiments.[31,33]

THE CASE OF HgTe - CdTe

Let us turn to a case where the effect of T should be expected to be larger and to be more clearly recognized. Such a case, where the offsets have also received much experimental attention, is the HgTe–CdTe heterojunction.[19,20,34,35] There is a large literature on this heterojunction due to the effort to develop infra-red devices based upon it.

The reader should note that, because HgTe is a semimetal, the antibonding-s Γ_6 level lies below the bonding-p Γ_8 level. However, by convention, one continues to label Γ_6 as "E_c" and Γ_8 as "E_v."

Kowalczyk et al. concluded from a room temperature optical experiment[34] that ΔE_v(HgTe/CdTe) = 0.35 ± 0.06 eV, i.e., that the HgTe valence band edge is higher than that of CdTe by this amount. This conclusion has been controversial because while others have repeated Kowalczyk et al.'s experiment,[36]

several other determinations of the band offset have implied values an order of magnitude lower (see the discussion in Ref. 19).

First we note that the experiments giving the smaller values for ΔE_v were all done at low temperatures (around 4K). Some of these estimated ΔE_v from the optical properties of superlattice structures. The methods used for those inferences are somewhat controversial. We choose to avoid that controversy by instead basing our discussion on estimates of ΔE_v derived entirely from electronic transport measurements.[37] Chow et al. deduced[35] $\Delta E_v(HgTe/CdTe) < 0.10$ eV at T = 4 K using the same structure that demonstrated[20] $\Delta E_v(HgTe/CdTe) = 0.35 \pm 0.06$ eV at T = 300 K, consistent with Ref. 34. In particular, they demonstrated negative differential resistance in a $Hg_{0.78}Cd_{0.22}Te/CdTe$ junction device[35] at T = 4K and argued forcefully that this observation requires $\Delta E_v(HgTe/CdTe) < 0.10$ eV.

We now consider whether the discrepancy re $\Delta E_v(HgTe/CdTe)$ can be ascribed to the expected temperature variation of the two E_v's. We first note that HgTe has the property that $\frac{d\Delta E_{cv}}{dT} > 0$ in sharp contrast to the $\frac{d\Delta E_{cv}}{dT} < 0$ behavior of CdTe, GaAs, AlAs, Si, and most other semiconductors. As shown by Heine and Van Vechten, this is because[23] thermal excitation of an electron-hole pair adds an electron in the Γ_8 bonding-p band and a hole in the Γ_6 antibonding-s band. Thus, the effect of thermally excited e_c's and h_v's is to stiffen, rather than to soften, the phonon modes and $\Delta S_{cv} < 0$. This means that for HgTe, and its alloys, E_v rises with increasing T while for CdTe, E_v falls. (See Fig. 3.) It then follows from Eq. (23) that if $E_v(HgTe) > E_v(CdTe)$ at T=0, then $\Delta E_v(HgTe/CdTe)$ will increase rapidly with T for T > 0. (If $E_v(HgTe) < E_v(CdTe)$ at T = 0, then with increasing T they would cross and

E_v(HgTe) will rapidly rise above E_v(CdTe).) The variation is particularly rapid because both terms in Eq. (23) are positive; there is no cancellation between them. Indeed, we expect

$$\Delta E_v(\text{HgTe/CdTe, T=300K}) - \Delta E_v(\text{HgTe/CdTe, T=0K}) = 0.77[(\Delta E_{cv}(\text{HgTe, T=300K}) - \Delta E_{cv}(\text{HgTe, T=0K})) + (\Delta E_{cv}(\text{CdTe, T=0K}) - \Delta E_{cv}(\text{CdTe, T=300K}))] \ . \quad (30)$$

To be quantitative, ΔE_{cv}(HgTe) increases[38] about 0.160 eV between 0 and 300K while ΔE_{cv}(CdTe) decreases[38] about 0.161 eV. Thus, the expected variation of ΔE_v is

$$\Delta E_v(\text{HgTe/CdTe, T=300K}) - \Delta E_v(\text{HgTe/CdTe, T=0K}) = 0.247 \text{ eV} \ . \quad (31)$$

To this can be added the effect of the alloy variation of E_v; Chow et al. set their lower limit E_v on the basis of a 78% Hg alloy rather than for pure HgTe. Approximating the variation of the E_v as linear, we estimate that if $\Delta E_v(T=0) = 0.10$ eV for the 78% alloy, as Chow et al. allow, then it is 0.13 eV for pure HgTe. (This should be a small overestimate due to the alloy disorder induced bowing of the band edges.[22,38]) We would then calculate

$$\Delta E_v(\text{HgTe/CdTe, T=300K}) = 0.13 + 0.25 = 0.38 \text{ eV} \ , \quad (32)$$

which is even larger than estimated by Kowalczyk et al. but within their error limit, $\Delta E_v(\text{HgTe/CdTe, T=300K}) = 0.35 \pm 0.06$ eV. Obviously, the value calculated at Eq. (32) can be reduced by allowing for the alloy bowing (an effect estimated to be roughly 0.01 eV) and by assuming $\Delta E_v(T=0)$ to be less than the maximum allowed by Chow et al. In addition, the variation of ΔE_{cv} (CdTe) may be less than reported in Ref. 38 and assumed here. Zanio

reported[39] the variation to be only 0.10 eV from 0 to 300K, which would reduce the estimate of Eq. (32) to 0.33 eV.

SUMMARY

We conclude that this variation with T is the likely explanation for the discrepancy between the experimental determinations of ΔE_v(HgTe/CdTe) by Chow et al. and by Kowalczyk et al.; both results can be quite correct. Furthermore, the very simple theory proposed here is consistent with all data available to us.

We have presented the thermodynamic perspective on the temperature dependence of the band offsets in heterojunctions. We have argued that the magnitude of the variation is directly proportional to the difference in the band gap temperature variation. The HgTe–CdTe and AlAs–GaAs systems were specifically discussed, but extensions to other materials combinations and to the effect of pressure on the band offsets follow in similar fashion.

ACKNOWLEDGEMENT

We are grateful to D.J. Wolford for very useful discussions of the AlAs–GaAs case. The work at Oregon State University was supported in part by the Air Force Office of Scientific Research under Contract No. AFOSR 86-0309.

FIGURE CAPTIONS

Fig. 1. The change in energy, etc., when an electron goes from a particular state at a particular point in the bulk of Ge to a particular state at a particular point in the bulk of GaAs cannot depend upon the path taken. Thus, the bulk-to-bulk band offset, the definition used by those who use the thermodynamic approach, cannot depend on the orientation of the junction and must be a transitive property.

Fig. 2. The empirical variation with temperature, as determined by Thurmond in Ref. 21, of the free energy, the enthalpy, and the entropy of the band gap, i.e., of the creation of pair of free carriers, in GaAs.

Fig. 3. Variation with T of the valence and conduction band edge free energies, E_v and E_c, for GaAs and AlAs.

Fig. 4. Variation with T of the valence and conduction band edge free energies for HgTe-CdTe. In this figure we have assumed ΔE_v (HgTe/CdTe) = 50 meV at T = 0 K and the variation with T of ΔE_{cv} for both HgTe and CdTe to be as reported in Ref. 38. The HgTe bands are labeled according to the prevailing "negative band gap" convention also used in Refs. 34-39. Thus, the antibonding-s Γ_6 level, which corresponds to the bottom of the conduction band in CdTe (and in GaAs), is labeled "E_c" even though it lies below the bonding-p Γ_8 level, which corresponds to the top of the valence band in CdTe (and in GaAs, AlAs, etc.), and which is labeled "E_v". The ΔE_v values quoted are indeed the differences of the Γ_8 levels across the junction.

REFERENCES

1) H. Kromer, Proc. Inst. Radio Eng. 45, 1535 (1957).

2) H.S. Rupprecht, J.M. Woodall, and G.D. Pettit, Appl. Phys. Lett. 11, 81 (1967).

3) Zh. Alferov, V.M. Andreev, V.I. Korol'kov, E.L. Portnoi, and D.N. Tret'yakov, Fiz. Tekh. Poluprovodni. 2, 1016 [Sov. Phys. Semicond. 2, 843] (1969).

4) M.B. Panish and S. Sumski, J. Phys. Chem. Solids 30, 129 (1969).

5) L. Esaki and R. Tsu, IBM J. Res. Dev. 14, 61 (1970).

6) R. Dingle, H.L. Stromer, A.C. Gossard, and W. Wiegmann, Appl. Phys. Lett. 33, 665 (1978).

7) H. Kromer, in Molecular Beam Epitaxy and Heterostructures, edited by L. L. Chang and K. Ploog (M. Nijhoff, The Hague, 1984).

8) J.R. Arthur, editor, Proceedings of the 3rd International MBE Conference, J. Vac. Sci. Technol. B 3, 1 (1985).

9) T.N. Casselman, editor, Proceedings of the 1984 U.S. Workshop on the Physics and Chemistry of Mercury Cadmium Telluride, J. Vac. Sci.Technol. A 3, 47 (1985).

10) R. Dingle, editor, Applications of Multiquantum Wells, Selective Doping, and Superlattices, (Academic Press, New York, 1987).

11) C. Herring and M.H. Nichols, Rev. Mod. Phys. 21, 185 (1949).

12) R.L. Anderson, Solid State Electron. 5, 341 (1962).

13) J.A. Van Vechten, Phys. Rev. 182, 891 (1969).

14) J. Tersoff, Phys. Rev. B 30, 4874 (1984).

15) J.A. Van Vechten, J. Vac. Sci. Technol. B 3, 1240 (1985).

16) M. Cardona and N.E. Christensen, Phys. Rev. B 35, 6182 (1987).

17) C.G. Van de Walle and R.M. Martin, Phys. Rev. B 35, 8154 (1987).

18) D.M. Bylander and L. Kleinman, Phys. Rev. Lett. 18, 2091 (1987).

19) J.-P. Faurie, C. Hsu, and T.-M. Duc, J. Vac. Sci. Technol. A 5, 3074 (1987).

20) D.H. Chow, J.O. McCaldin, A.R. Bonnefoi, T.C. McGill, I.K. Sou, and J.-P. Faurie, Appl. Phys. Lett. 51, 2230 (1987).

21) C.D. Thurmond, J. Electrochem. Soc. 122, 1133 (1975).

22) J.A. Van Vechten in Handbook on Semiconductors, Vol. 3, Materials, Properties and Preparations, edited by S. P. Keller (North Holland, Amsterdam, 1980) chapt. 1 p. 1.

23) V. Heine and J.A. Van Vechten, Phys. Rev. B 13, 1622 (1976).

24) R.M. Martin, Phys. Rev. 186, 871 (1969).

25) W. Weber, Phys. Rev. Lett. 33, 371 (2974).

26) J.A. Van Vechten, Phys. Rev. B 31, 2508 (1985).

27) L. Vina, H. Hochst, and M. Cardona, Phys. Rev. B 31, 958 (1985).

28) V. Heine and C.H. Henry, Phys. Rev. B 11, 3795 (1975).

29) J.A. Van Vechten and C.D. Thurmond, Phys. Rev. B 14, 3539 (1976).

30) J.A. Van Vechten, Phys. Rev. B 13, 946 (1976); Phys. Rev. B 14, 1781 (1976).

31) D.J. Wolford, T.F. Keuch, and J.A. Bradley, J. Vac. Sci. Technol. B 4, 1043 (1986).

32) F.M. Vorobkalo, K.D. Glinchuk, and V.F. Kovalenko, Fiz. Tekh. Poluprovodni 9, 998 [Sov. Phys. Semicond. 9, 656] (1975).

33) J. Batey and S.L. Wright, J. Appl. Phys. 59, 200 (1986).

34) S.P. Kowalczyk, J.T. Cheung, E.A. Kraut, and R.W. Grant, Phys. Rev. Lett. 56, 1605 (1986).

35) D.H. Chow, T.C. McGill, I.K. Sou, J.P. Faurie, and C.W. Nieh, Appl. Phys. Lett. 52, 54 (1988).

36) C.K. Shih and W.E. Spicer, Phys. Rev. Lett. 58, 2594 (1987).

37) K.J. Malloy and J.A. Van Vechten, to be published.

38) G.L. Hansen, J.L. Schmidt, and T.N. Casselman, J. Appl. Phys. 53, 7099 (1982).

39) K. Zanio in Semiconductors and Semimetals, edited by R.K. Willardson and A.C. Beer (Academic Press, New York, 1978), p. 99.

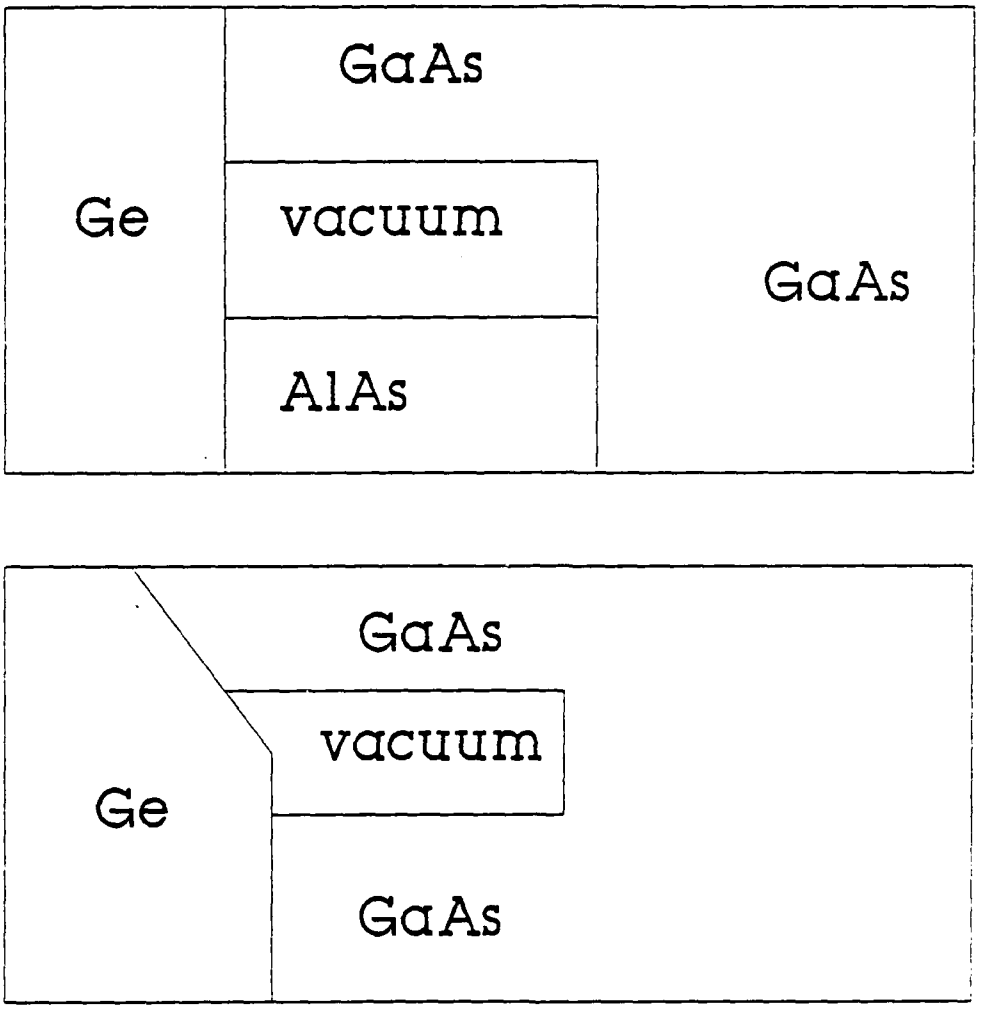
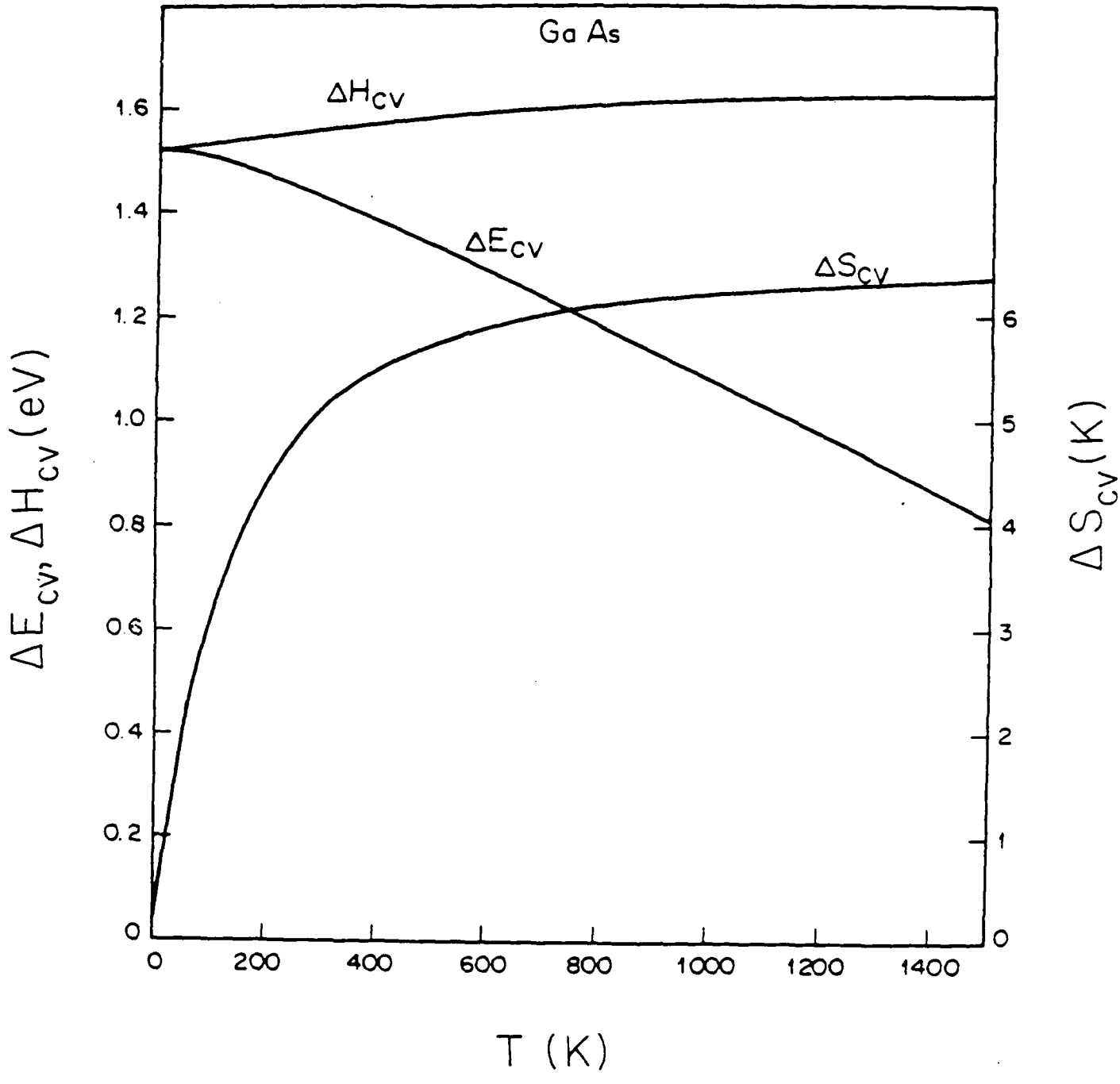
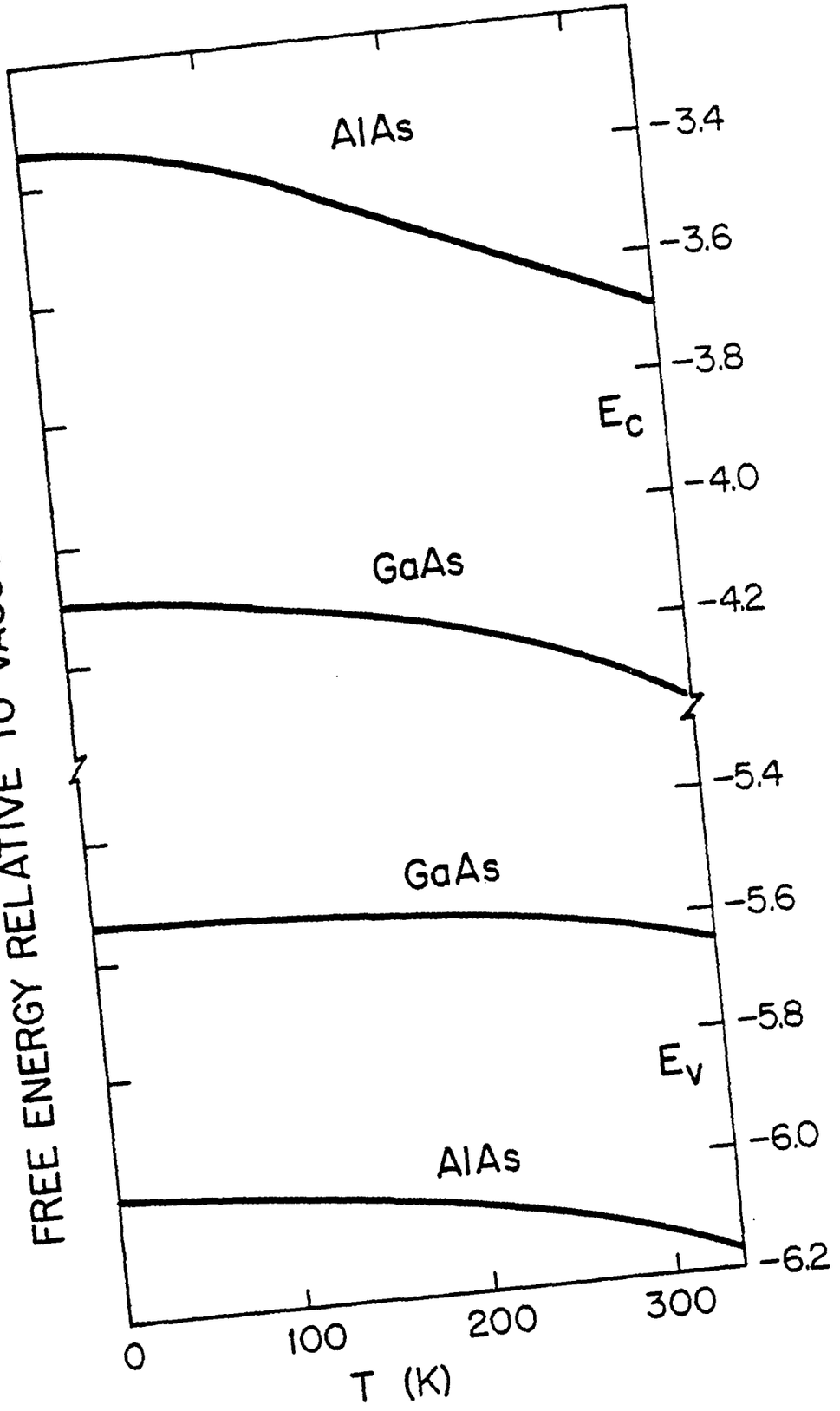
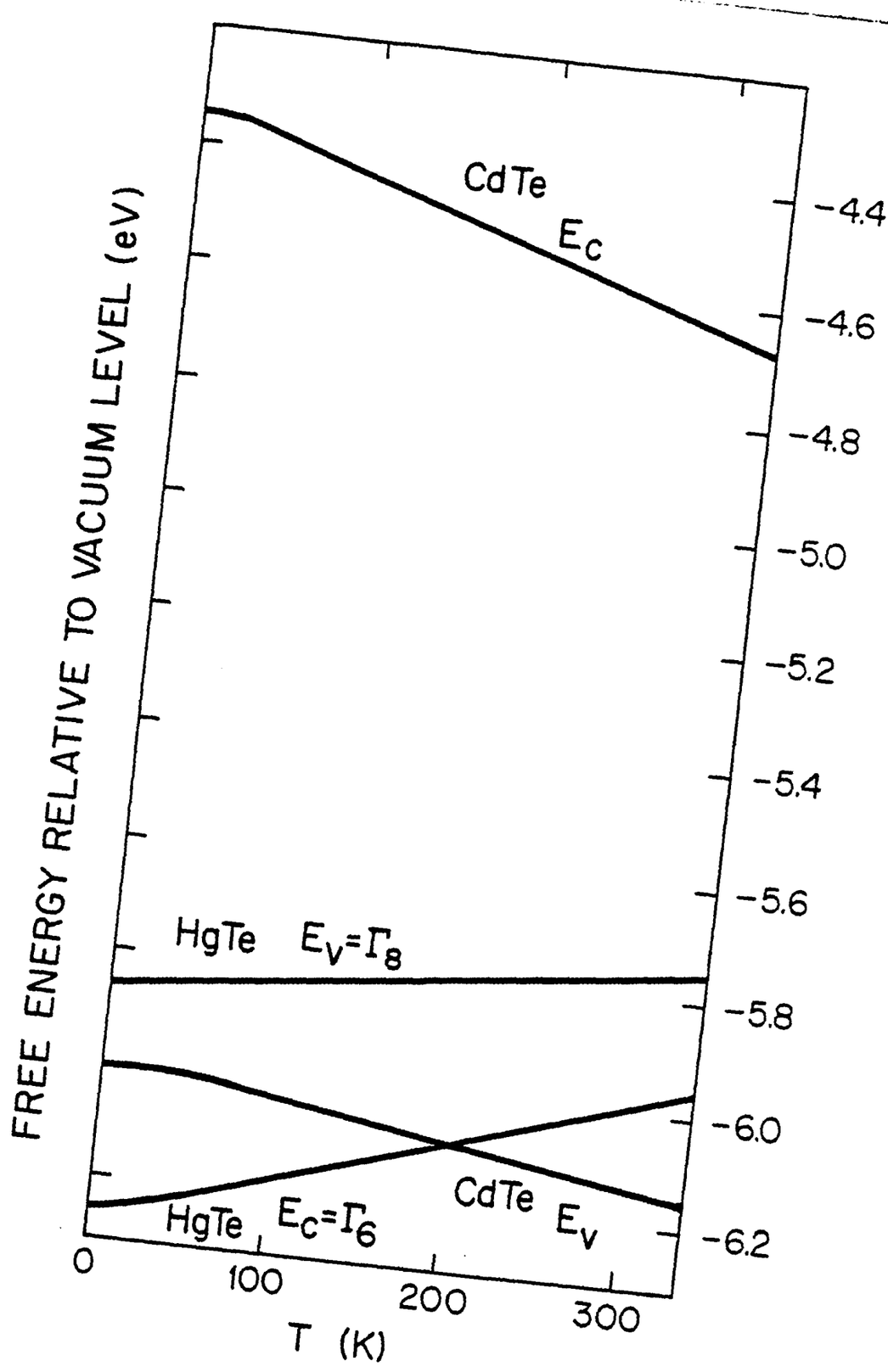


Fig. 1


Ga As
ΔH_{CV}
ΔE_{CV}
ΔS_{CV}
$\Delta E_{CV}, \Delta H_{CV}$ (eV)
ΔS_{CV} (K)
T (K)
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1
2
3
4
5
6
0
200
400
600
800
1000
1200
1400


AlAs
GaAs
GaAs
AlAs
E_c
E_v
-3.4
-3.6
-3.8
-4.0
-4.2
-5.4
-5.6
-5.8
-6.0
-6.2
0
100
200
300
T (K)
FREE ENERGY RELATIVE TO VACUUM LEVEL (eV)


FREE ENERGY RELATIVE TO VACUUM LEVEL (eV)
-4.4
-4.6
-4.8
-5.0
-5.2
-5.4
-5.6
-5.8
-6.0
-6.2
CdTe
E_c
HgTe $E_V=\Gamma_8$
HgTe $E_c=\Gamma_6$
CdTe E_V
0
100
200
300
T (K)

Vacancy First and Second Neighbor Hopping at a Compound Semiconductor Interface - Insights from Computer Simulation

J.A. Van Vechten and U. Schmid*

Center for Advanced Materials Research

Department of Electrical and Computer Engineering

Oregon State University, Corvallis OR 97331-3202

ABSTRACT

We have now demonstrated that the ratio of the activation enthalpies for nearest neighbor hopping of the two types of vacancy in InP are in the 4 to 1 ratio of the masses of the atoms that hop and are in good agreement with the values predicted by the Ballistic Model, BM, which identifies the activation enthalpy as a kinetic energy. A consequence is that relatively light atoms should hop to second neighbor vacancy sites as easily as heavy atoms hop to nearest neighbor vacancies, and do so without the complication of forming or annihilating antisites. E.g., in GaAs B, C, N, Al, Si and P are predicted to hop to second neighbor sites as easily as As hops to a first neighbor site. Of course, they hop to nearest neighbor sites even more readily. The import of this on the array of point defects predicted to occur as a heterojunction anneals is rather complex. Experiment supports our contention of complexity. E.g., in As rich, n-type AlGaAs Mei et al. report D(Al) α n^1 exp(-2.9eV/kT) if the doping is with Te but D(Al) α n^3 exp(-4eV/kT) if the doping is with Si, where n is the free electron density.We have studied the Al diffusion in AlGaAs with the aid of a microcomputer simulation ex consequences of ones assumptions re thermochemical c level. Our previously published BM assumptions le t Si

Jan
Van Vechten

promotes the conversion of Ga vacancies to trivacancies while Te does not; they also identify 2.9 eV as the sum of the Ga second neighbor hopping enthalpy and the enthalpy of formation of single negative Ga vacancy in n-type GaAs and identify 4 eV as the sum of the enthalpy of formation of the trivacancy plus its nearest neighbor hopping energy in GaAs. The simulation also finds that Al by itself converts As vacancies to trivacancies plus metal antisite. It shows the trivacancies to diffuse very rapidly and to entrain Al. We propose that the cation sublattice dominance of Si induced host interdiffusion follows from this entrainment. We also suggest these complexes are related to DX and account for the anomalous variation of the parameters of that defect with Al concentration.